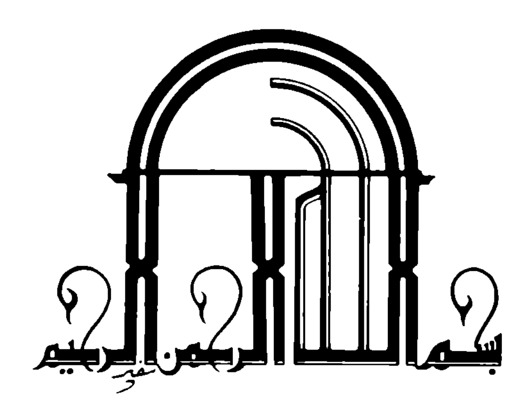

مكتبة الملك فهد الوطنية
الرياض
١٤١٠
مطبوعات
مكتبة الملك فهد الوطنية
السلسلة الثانية
(٦٤)
الويب
نظام استرجاع المعلومات الكوني
سيد ربيع سيد إبراهيم
الرياض
١٤٣٠هـ / ٢٠٠٩م



الـويـــــــب :

نظام استرجاع المعلومات الكوني

مطبوعات
مكتبة الملك فهد الوطنية
السلسلة الثانية
(٦٤)

تعنى هذه السلسلة بنشر الدراسات والبحوث
في إطار علم المكتبات والمعلومات بشكل عام

الويـــــب :

نظام استرجاع المعلومات الكوني

THE WEB :
The Global Information Retrieval System

تاليف

سيد ربيع سيد إبراهيم

مكتبة الملك فهد الوطنية

الرياض : ١٤٣٠هـ/ ٢٠٠٩م

ح مكتبة الملك فهد الوطنية ، ١٤٣٠ هـ

فهرسة مكتبة الملك فهد الوطنية أثناء النشر

إبراهيم ، سيد ربيع سيد

الويب : نظام استرجاع المعلومات الكوني./ سيد ربيع سيد
إبراهيم.- الرياض، ١٤٣٠هـ.

٣٧٨ ص ، ١٧ × ٢٤ سم.- (الثانية ؛ ٦٤)

ردمك : ٤ ـ ٣٢٤ ـ ٠٠ ـ ٩٩٦٠ ـ ٩٧٨

١ . الإنترنت أ . العنوان ب . السلسلة

ديوي ٠٠٤,٦٧ ١٤٣٠/١٦٩٥

رقم الإيداع : ١٤٣٠/١٦٩٥

ردمك : ٤ ـ ٣٢٤ ـ ٠٠ ـ ٩٩٦٠ ـ ٩٧٨

ص.ب : ٧٥٧٢
الرياض: ١١٤٧٢ المملكة العربية السعودية
هاتف: ٤٦٢٤٨٨٨
فاكس: ٤٦٤٥٣٤١

إهداء

إلى أناس أحببتهم .. أسعدوني وتأثرت بهم ..

فمنهم من غمرني بعلمه .. ومنهم من أسرني بفضله ..

ومنهم من هم عمري .. وأنا أملهم

أستاذي د. هاني محيي الدين ، أستاذتي د. فيدان عمر

والداي وزوجتي ودارين

بسم الله الرحمن الرحيم

﴿ وَعَلَّمَكَ مَا لَمْ تَكُن تَعْلَمُ ۚ وَكَانَ فَضْلُ ٱللَّهِ عَلَيْكَ عَظِيمًا ﴾

النساء: ١١٣ صدق الله العظيم

المحتويات

قائمة الجداول

قائمة الأشكال



مقدمة الدراسة :

أولاً : التمهيد:

استخدم الإنسان المعلومات المسجلة لتناقل الأحداث وتبادل المعرفة عبر الأزمنة والأماكن المختلفة. وقد شكلت المعلومات المسجلة في أماكن حفظها واختزانها نظمًا لتنظيم وبحث المعلومات. وعلى ذلك يمكن القول إن مصطلح نظام استرجاع المعلومات لا يرتبط بالحضارة الرقمية التي يعيشها الإنسان الآن، وإنما يرتبط بمختلف أشكال تسجيل وحفظ وتدوين المعلومات منذ بداية الكتابة على الجدران وحتى استخدام اللغة الثنائية 010101 لغة الحاسب الآلي. واستخدام مصطلح نظام استرجاع المعلومات إنما يدل على مختلف مؤسسات ومرافق المعلومات التقليدية القديمة والإلكترونية الحديثة، فمكتبة الإسكندرية القديمة التي حوت آلاف البرديات هي نظام استرجاع للمعلومات، ومكتبات المساجد في القرون الإسلامية الأولى هي نظم استرجاع للمعلومات، والمكتبات النوعية التقليدية هي نظم استرجاع للمعلومات، ثم قواعد البيانات على الأقراص المدمجة ومحركات البحث على الويب، بل والويب ذاتها هي نظام لاسترجاع المعلومات الرقمية.

إن مصطلح نظم استرجاع المعلومات يدل على مفهوم عام ينسحب على مختلف مرافق المعلومات على اختلاف أنواعها ودرجاتها. غير أن الإنتاج الفكري الأجنبي والعربي دائمًا ما يتحدث عن نظم استرجاع المعلومات من منظور واحد فقط، ألا وهو استراتيجيات ولغات بحث المعلومات. وقد

استعرض الباحث الإنتاج الفكري في قاعدة مستخلصات علم المكتبات والمعلومات LISA الخاص بمفهوم نظم استرجاع المعلومات فلم يجد من الدراسات ما أعطى لهذا المفهوم حقه كاملا. ولم يرد مفهوم نظم استرجاع المعلومات في قاموس مصطلحات المكتبات وعلم المعلومات على الخط المباشر ODLIS بشكل مخصص ؛ حيث عرف ODLIS مفهوم استرجاع المعلومات على أنه " العملية أو الإجراء أو الطريقة التي تنتقي مجموعة من المعلومات المسجلة لاستدعائها بحسب مجموعة من المصطلحات البحثية التي تعبر عن حاجة موضوعية محددة ، وهذه العملية قد تكون تقليدية كما يتم في استخدام الفهارس البطاقية في المكتبات أو قد تكون إلكترونية كما يتم في الفهارس على الخط المباشر OPAC أو مراصد البيانات الببليوجرافية Bibliograghic databases." ومن ثم فإن هذا المفهوم يركز على الجانب الإجرائي والأدائي الذي تتم عليه عملية استرجاع المعلومات سواء أكان ذلك في البيئة التقليدية أو الرقمية.

إن ظهور الإنترنت والويب وأدوات بحث مصادرها الرقمية المختلفة الأشكال والمتباينة الوسائط والمتعددة الموضوعات قد دفع الباحث إلى النظر لما أحدثته البيئة الرقمية المتمثلة في الإنترنت وخاصة الويب في تغيير إدارة وعمل ومهام نظم استرجاع المعلومات. فلقد ظلت نظم استرجاع المعلومات حتى الآلية منها تسير سيرًا منطقيًا في إجراءات تشغيلها ومعالجتها وتنظيمها للمعلومات ثم استراتيجيات بحثها مرة أخرى، غير أن الويب قد أحدثت تغييرات كثيرة في ملامح نظم استرجاع المعلومات، وإن جاز لنا هنا أن نطلق عليها نظام النظم أو

نظام نظم استرجاع المعلومات، وبدت هذه التغييرات كظاهرة جديدة تتعلق بتنظيم ومعالجة واسترجاع المعلومات الرقمية. وقد أراد الباحث من خلال هذا البحث دراسة فضاء الويب ومعايشة عناصر بنيتها وملامح تشغيلها كنظام كوني لاسترجاع المعلومات. والهدف من ذلك كله هو تقديم رؤية جديدة للتعامل مع الويب بيسر وتحقيق الكفاءة المرجوة في استرجاع المصادر الرقمية على الويب.

ثانيًا : مشكلة الدراسة:

تكمن مشكلة الدراسة في زيادة تدفق المعلومات إلى خادمات الويب حتى وصل مقدار اختزانها إلى ٧٥٠٠ تيرا بايت من المعلومات الرقمية، وهذه المعلومات توجد في أشكال مختلفة مثل صفحات ومواقع وقواعد بيانات الويب، كما أنها توجد في وسائط مختلفة كالنص والصوت والصورة الثابتة والمتحركة، إلى جانب تعدد موضوعاتها. ومن عناصر المشكلة البحثية أيضًا اختلاف أدوات بحث الويب في المعالجة والتشغيل لاسترجاع معلومات الويب ؛ فتعمل محركات البحث باستخدام مجموعة من البرامج الفرعية ذاتية العمل وتختص بصفحات ومواقع الويب، على الجانب الآخر، فإن أدلة بحث الويب تعتمد على العنصر البشري في إجراء عمليات المعالجة والتشغيل لدليل البحث.

إن مشكلة الويب لم تكمن فقط في أعداد مصادر المعلومات سريعة التنامي، أو صعوبة المعالجة والاسترجاع داخل أدوات بحث الويب، وإنما تمثلت مشكلة الويب أيضًا في ظهور مجموعة مستحدثة من أشكال المعلومات مثل ملفات الصور والصوت ومواد الرسوم الحيوية، التي تحتاج بدورها إلى مجموعة

من المعايير لضبط عمليات المعالجة الفنية وتنظيم المحتوى. أيضًا مما يزيد المشكلة البحثية تعقدًا هو ظهور نوع جديد من نظم استرجاع المعلومات ليس لاسترجاع المعلومات الببليوجرافية أو المعلومات الأولية، وإنما لاسترجاع نظم الاسترجاع ذاتها، خير ما يمثلها هو نظم استرجاع قواعد بيانات الويب التي يطلق عليها الويب غير المرئية أو الخفية invisible web databases، وهو ما أفرد الباحث له دراسة الدكتوراة. ويمكن القول إن الويب في بنيتها كنظام استرجاع للمعلومات أوجدت مشكلة في أن نقصر مفهوم استرجاع المعلومات على ما كان عليه قبل ظهور الإنترنت والوقوع في بيت العنكبوت (الويب).

ثالثًا : تساؤلات الدراسة:

تسعى الدراسة للإجابة على التساؤلات التالية:

١- ما التطور التاريخي الذي مرت به نظم استرجاع المعلومات؟

٢- ما المفاهيم التي بني عليها نظام استرجاع المعلومات في عمليات الإضافة والمعالجة والاسترجاع؟

٣- كيف يمكن تغيير منظورنا لمفهوم نظم استرجاع المعلومات في ضوء ما أحدثته الويب؟

٤- كيف يمكن النظر إلى الويب على أنها نظام استرجاع للمعلومات الرقمية؟

٥- ما خصائص عمل الويب كنظام كوني لاسترجاع المعلومات الرقمية؟

٦- ما طبيعة الخصائص التفاعلية والديناميكية التي تتسم بها الويب كنظام استرجاع؟

رابعًا : أهمية الدراسة:

تأخذ هذه الدراسة أهميتها واختلافها عن الدراسات الأخرى التي تحدثت عن نظم الاسترجاع، انطلاقًا من العناصر التالية:

١- تمثل نظم استرجاع الويب شكلاً مختلفًا عن غيرها من أنواع نظم الاسترجاع الأخرى، لما تتسم به من خصوصية في عمليات المعالجة والتنظيم وبحث واسترجاع المعلومات. فضلاً عن خصائص أخرى مثل التحديث والتغيير المتلاحق لمصادر معلومات الويب.

٢- لم يتطرق الإنتاج الفكري العربي من قبل إلى الويب على أنها نظام متكامل يبدأ عند تصميم وإضافة المعلومات الرقمية في فضائه، ثم معالجتها وتطوير سبل استرجاعها مرة أخرى.

٣- تركز الدراسة الضوء على التحول والتبديل في المفاهيم وجوانب التنظير المختلفة التي تبنى عليها نظم استرجاع المعلومات.

٤- تغطي الدراسة بعضًا من الخصائص التي تتسم بها نظم استرجاع المعلومات العاملة في بيئة الويب، وهي خصائص مستحدثة ارتبطت بالنظر للويب على أنها النظام الكوني لاسترجاع المعلومات.

٥- عملت الدراسة على التغطية الشاملة لمختلف جوانب التقييم وقياس الكفاءة داخل نظم استرجاع المعلومات، إضافة إلى إرساء ملامح التقييم واختلافها بين كل من النظم التقليدية ونظم الخط المباشر من ناحية، ونظم استرجاع معلومات الويب من ناحية أخرى.

٦- تمثل الدراسة بداية الحديث عن الويب كنموذج جديد لنظم استرجاع المعلومات، من أهم سماته التفاعل والتحديث والتغيير المتلاحق، فضلاً عن حداثة أساليب المعالجة والتنظيم ولغات الحوار مع المستفيدين.

خامسًا : فروض الدراسة:

١- أحدثت الويب كنظام كوني للمعلومات الرقمية تغييرًا في مفاهيم وأداء وبناء نظم استرجاع المعلومات.

٢- إن التغيير الذي أحدثته الويب في مفاهيم وإجراءات نظم استرجاع المعلومات قد صنع جيلاً مختلفًا من نظم الاسترجاع له خصائصه وملامحه التي يختلف بها عن غيره من أجيال نظم استرجاع المعلومات.

سادسًا : أهداف الدراسة :

تسعى الدراسة إلى تحقيق الأهداف التالية:

١- التعرف إلى بيئة الويب والنظر إليها كنظام عالمي لاختزان واسترجاع المعلومات الرقمية.

٢- رصد أهم ملامح وخصائص العمل والبناء في نظام استرجاع معلومات الويب.

٣- رصد التغييرات في المفاهيم والأداء التي طرأت على نظم استرجاع المعلومات.

٤- التعرف إلى التطور التاريخي والملامح الزمنية التي رسمت على وجه نظم استرجاع المعلومات.

٥- دعم عمليات التحول والأخذ بأسباب الرقمنة؛ لتطوير أداء وبنية نظم استرجاع المعلومات.

٦- التعرف إلى بعض خصائص المعلومات التي تؤثر في أنماط نظم الاسترجاع مثل التفاعلية والديناميكية التي تكون عليها هذه المعلومات.

سابعًا : منهج الدراسة :

استخدم الباحث المنهج الوصفي التحليلي لاستقراء الويب كظاهرة تمثل تطورًا لنظم استرجاع المعلومات، واعتمد الباحث على متابعة واستقراء الإنتاج الفكري حول هذه الظاهرة، ثم تحليله وتأويله في تابع فصول الدراسة.

أدوات البحث :

استخدم الباحث أداة البحث الوثائقي المتمثلة في بحث الإنتاج الفكري ومصادر المعلومات على الويب. ومن بين مصادر جمع الإنتاج الفكري التي اعتمد عليها الباحث ما يلي:

1. www.magportal.com
2. www.google.com
3. www.yahoo.com
4. www.informationr.net
5. www.findarticles.com

ثامنًا : حدود الدراسة :

سار الباحث في بحث وتحليل الإنتاج الفكري حول الويب ونظم استرجاع المعلومات وفقًا لحدود البحث التالية:

❖ الحدود الموضوعية :

يتعرض الباحث في دراسته لموضوع نظم استرجاع المعلومات وتناول شبكة الويب كنموذج عالمي لاسترجاع المعلومات الرقمية، وكجزء من الشبكة

العالمية. وتناول نظم استرجاع المعلومات بما ينطوي على الجانب التاريخي وتأثير الويب في تطوير نظم الاسترجاع مفهوماً وأداءً.

❖ **الحدود النوعية :**

تتناول الدراسة نظم استرجاع الويب من ناحية المفاهيم والبناء والمكونات لبيئتي نظم استرجاع أوعية المعلومات التقليدية والمعلومات الرقمية على الويب؛ بالإضافة إلى التعرض لمختلفة نماذج نظم استرجاع المعلومات الداعمة لاسترجاع النصوص والوسائط المتعددة.

❖ **الحدود الشكلية :**

تلقي الدراسة الضوء على نظم استرجاع معلومات الويب وتناول مصادر المعلومات ذات الشكل الرقمي.

❖ **الحدود الزمنية:**

تعالج الدراسة موضوع نظم استرجاع المعلومات ومراحل تطورها السابقة منذ بداية الأربعينات، ثم التوقف الزمني عند السنوات الأُوَل من هذه الألفية؛ للتركيز على خصائص بيئة الويب كنظام كوني لاسترجاع المعلومات الرقمية.

❖ **الحدود المكانية :**

تتناول الدراسة نظم استرجاع المعلومات الرقمية على الويب كنظام استرجاع ومستودع لمعلومات العالم الرقمية.

❖ **الحدود اللغوية :**

تتناول الدراسة نظم استرجاع المعلومات التي تستخدم اللغة الإنجليزية كلغة أم لمستودع المعلومات الكوني، بجانب الإشارة لآليات اللغة العربية.

تاسعًا : الدراسات السابقة:

سبقت هذه الدراسة الكثير من الدراسات التي تناولت موضوع نظم استرجاع المعلومات سواء التقليدية أو الرقمية. غير أن دراستنا هذه قد تناولت ملامح جيل جديد من نظم استرجاع المعلومات تأثر بظهور عملاق هذه النظم، ألا وهو الويب كنظام استرجاع للمعلومات الرقمية يشمل العالم أجمع. وقد تأثرت المفاهيم وجوانب تناول وقياس كفاءة نظم استرجاع المعلومات وفقًا للملامح التي تتسم بها الويب، وأهم هذه الملامح هي التغيير المتسارع والتفاعلية العالية والديناميكية القوية التي تتميز بها آليات تخليق وتنظيم وبحث واسترجاع محتوى الويب الرقمي. ويمكن عرض بعضٍ من الدراسات التي تناولت نظم استرجاع المعلومات وجوانب الاختلاف في هذه الدراسة.

الدراسة الأولى :

لانكستر، ولفرد. نظم استرجاع المعلومات[1].

حصر لانكستر في دراسته مختلف الجوانب المتعلقة بنظم المضاهاة، واعتبر مراحل تطور المضاهاة هي مراحل تطور نظم استرجاع المعلومات، مما يعكس الأهمية التي تتسم بها عملية المضاهاة ويتوقف عليها أداء نظام الاسترجاع. اهتم لانكستر أيضًا بطبيعة الاحتياجات الموضوعية للمستفيدين وصور التعبير عن حاجاتهم من الحاجة الكامنة إلى الحاجة المصورة، كما أورد أيضًا بعضًا من معاملات قياس كفاءة نظم الاسترجاع.

---

(١) لانكستر، ولفرد. **نظم استرجاع المعلومات** ؛ ترجمة حشمت قاسم .- القاهرة : دار غريب، ١٩٨١م.

إن المرحلة الزمنية التي تحدثت عنها دراسة لانكستر تختلف في شكلها ومضمونها عن عصر نظم استرجاع الويب الذي نشهده، مما يعني أن هناك من المستحدثات التي طرأت على نظم الاسترجاع من ناحية المفهوم والأداء والتقييم. ومن ثم كانت دراستنا هذه التي تعكس المتغيرات الآنية لنظم استرجاع المعلومات.

الدراسة الثانية :

فاتن سعيد بامفلح. أساسيات نظم استرجاع المعلومات الإلكترونية[١].

اتسمت هذه الدراسة بالشمولية في عرض جوانب نظم استرجاع المعلومات المختلفة ؛ حيث تناولت أسس بناء نظم استرجاع المعلومات وبعضًا من هذه النظم على الإنترنت، ثم التركيز على معاملات الاستدعاء والتحقيق في أثناء التعرض لتقييم نظم الاسترجاع. غير أن دراستنا تختلف عن ذلك في بعضٍ من الجوانب أهمها :

- إلقاء الضوء على ما أحدثته الويب من تغييرات في مفاهيم الأنشطة والإجراءات داخل نظم استرجاع المعلومات.
- التعامل مع معاملات التقييم بشمولية أكثر تصل إلى تحليل بعض من الجوانب العامة مثل جوانب تختص بالنظام ذاته، وجوانب تخص عمليات المعالجة الفنية، وجوانب تخص النتائج المسترجعة. فضلا عن الإشارة لنماذج قياس كفاءة نظم استرجاع المعلومات الثابتة والديناميكية أو المتغيرة.

---

(١) فاتن سعيد بامفلح. **أساسيات نظم استرجاع المعلومات الإلكترونية**.- الرياض : مكتبة الملك فهد الوطنية، ٢٠٠٧م
http://www.kfnl.gov.sa/idarat/alnsherel/asaseiatnzmstrgaaalmalomat/pdf/ f1.pdf

- التركيز عن الويب كنظام لاسترجاع المعلومات ودراسة خصائصها في التنظيم والمعالجة والاسترجاع، وكيفية اختلافها عن بيئات نظم استرجاع المعلومات الأخرى.
- توضيح الجوانب المختلفة للتعامل مع لغات المعالجة والاسترجاع في نظم المعلومات ؛ حيث تعرضت الدراسة لمفهوم اللغة داخل نظم الاسترجاع، والفارق بين اللغتين الطبيعية والمقيدة في عمليات التنظيم والبحث. ثم التأكيد على أهم لغات وأساليب بحث المعلومات في البيئة الرقمية.

الدراسة الثالثة :

Chowdhury , G. G. Introduction to Modern Information Retrieval[١]

تتناول الدراسة المفاهيم والمكونات والوظائف داخل نظم استرجاع المعلومات، إلى جانب دراسة طبيعة قواعد البيانات وأنواعها. ثم تعرض للتحليل الموضوعي لمواد المعلومات والوصف الفني لها ، بجانب التركيز على نظم الاسترجاع على الأقراص الضوئية. واسترجاع الأنواع المختلفة من المعلومات مثل الوسائط المتعددة. وتختلف دراستنا الحالية عن ذلك في أنها تركز على نظم استرجاع المعلومات الأكثر ديناميكية وتغييرًا من نظم الأقراص الضوئية أو على الخط المباشر؛ إضافة إلى تخصيص الحديث عن نظم المعلومات المصاغة بلغات الترميز المعيارية مثل HTML AND XML.

---

(١) Chowdhury , G. G. Introduction to Modern Information Retrieval .- 2nd ed .- London: Facet publishing , 2003.

مأخوذة عن دراسة فاتن سعيد بامفلح. **أساسيات نظم استرجاع المعلومات الإلكترونية.**

الدراسة الرابعة :

White, Ryen W. Implicit Feedback for Interactive Information Retrieval. (1)

ركزت هذه الدراسة على سبل دعم التفاعل بين المستفيد والنظام من خلال دراسة بعض الأساليب الخاصة بالتلقيم المرتد أو التغذية المرتدة المرتبطة بالنتائج Relevance feedback (RF) technique. حيث طرح الباحث بعضًا من الواجهات التي ترتب نتائج البحث بأشكال مختلفة اعتمادًا على آليات إحصائية وهرمية لترتيب النتائج المسترجعة، ويستطيع المستفيد استخدام هذه الآليات لطرح وجهات نظره في كفاءة نظام الاسترجاع في استدعاء الوثائق المرتبطة بموضوع البحث. ومن ثم تختلف هذه الأطروحة عن دراستنا هذه في تناول نظم استرجاع المعلومات من جانب التقييم وقياس الكفاءة فقط، خلاف ما نهدف إليه من تناول نظم استرجاع المعلومات الرقمية مفهومًا وبنيةً وأداءً ثم تناول نظم استرجاع الويب من أكثر من جانب، بالإضافة إلى الحديث عن تقييم النظم وقياس الكفاءة من خلال معايير وعناصر تقييم شاملة.

عاشرًا : محتوى الفصول:

احتوت الدراسة على ستة فصول شكلت جميعًا الجوانب المختلفة لنظم الاسترجاع الرقمية مع التركيز على الويب بوصفه نظامًا كونيًا لاسترجاع المعلومات. وجاءت الفصول ومحتوياتها كالتالي:

---

(1) White, Ryen W. Implicit Feedback for Interactive Information Retrieval, University of Glasgow, 2004, (doctoral thesis), cited at https://research.microsoft.com/~ryenw/thesis/thesis- files/RyenWhiteThesis.pdf.

الفصل الأول : تضمن الفصل الأول الحديث عن مفهوم نظم استرجاع المعلومات ثم مكونات وبنية نظم الاسترجاع، التي تتكون من المدخلات والمخرجات والعمليات الداخلية. وتم التركيز في هذا الفصل على عمليات التحليل الشكلي والموضوعي في البيئات الرقمية، وإلقاء الضوء على أهم المعايير المستخدمة في ذلك.

الفصل الثاني : اهتم الفصل الثاني بالعوامل التي لا يمكن لنظام استرجاع المعلومات البناء أو أداء مهامه إذا ما غاب أحدها. وهذه العوامل تمثلت في اللغة التي يتحاور بها النظام مع المستفيد، والوثيقة التي تمثل أساس بناء نظام المعلومات، والمستفيد الذي يمثل هدف نظم استرجاع المعلومات منذ بدايته للوصول إلى تلبية احتياجاته من المعلومات. وبجانب ذلك فقد تضمن الفصل تقسيمًا مقترحًا لأنواع نظم استرجاع المعلومات اعتمادًا على بيئات وخصائص ووثائق نظم استرجاع المعلومات.

الفصل الثالث : ركز هذا الفصل على اثنين من أهم الملامح التي تتسم بها الويب وهما عنصرا التفاعلية بين جوانب النظام وبعضها البعض وبين المستفيد والنظام، وعنصر الديناميكية المتمثل في حركة ودوران المحتوى الرقمي داخل مستودعات الويب، والمتمثل أيضًا في أداء نظم استرجاع الويب، وتحوره للوصول إلى حد إرضاء المستفيدين بإتاحة المحتوى الرقمي المرضي. وتم تطبيق عنصري التفاعلية والديناميكية على الواقع الذي تعمل به نظم استرجاع الويب.

الفصل الرابع : عني الفصل الرابع بدراسة الويب بوصفه نظام استرجاع كونيًا للمعلومات، وقد تمثل ذلك في دراسة خصائص بنية المعلومات وأنماط التنظيم لمحتوى الويب الذي انعكس على واقع تنظيم مواقع وصفحات وقواعد

بيانات الويب. ثم تطرق الفصل إلى خصائص بيئتي نظام الويب المرئية وغير المرئية في استرجاع المعلومات الرقمية.

الفصل الخامس : يمثل هذا الفصل الجانب التطبيقي من الدراسة ، حيث ركز على بعض نماذج نظم استرجاع المعلومات والوسائط المتعددة التي توجد على الويب ؛ وذلك في إطار الحديث عن طبيعة المعلومات التي تحتويها هذه النظم وسبل تنظيمها واسترجاعها في البيئة الرقمية.

الفصل السادس : عرض الباحث في الفصل الأخير من الدراسة طرحًا لإطار شامل لتقييم وقياس كفاءة نظم استرجاع المعلومات ؛ وذلك من خلال الوقوف على الجوانب العامة التي يقاس من خلالها أداء النظام ثم العناصر الفرعية المكونة للإطار العام. وكانت الجوانب العامة للتقييم وقياس الكفاءة تتمثل في جوانب تخص النظام ذاته ، وجوانب تخص المستفيد ، وجوانب تخص النتائج المسترجعة ، وجوانب تخص العمليات الداخلية للنظام. ثم تم تناول كل واحد من هذه الجوانب مع تتابع عناصره بالتفصيل.

حادي عشر : مفاهيم الدراسة :

تتبنى الدراسة مجموعة من المفاهيم للمصطلحات التالية :

١- **نظم استرجاع المعلومات :** تعرفها الدراسة على أنها " هي النظم التقليدية أو الرقمية المكونة من مجموعة من العناصر أو البرامج الفرعية التي تعمل على إضافة واختزان وتنظيم ومعالجة المعلومات حتى يتم استرجاعها مرة أخرى وفقًا للحاجة الموضوعية للمستفيدين."

٢- **المعالجة الفنية :** هي مجموعة العمليات الفنية الداخلية التي يقوم بها نظام استرجاع المعلومات لوصف وتحليل محتوى الوثائق التي يضمها النظام.

٣- **الوصف الفني** : عمليات الوصف الشكلي أو المادي لمصادر المعلومات سواء التقليدية أو الرقمية.

٤- **الميتاداتا:** مجموعة البيانات أو الحقول أو عناصر الوصف لمصادر المعلومات الرقمية، ويطلق عليها البيانات الخلفية أو البيانات عن البيانات.

٥- **التنظيم** : مجموعة عمليات تحليل المحتوى للوثائق داخل نظام الاسترجاع، بهدف استخراج الكلمات الدالة على المحتوى، ويتم ذلك ضمن عملية التكشيف لوثائق النظام.

٦- **الويب غير المرئية:** هي مجموعة مواقع الويب التفاعلية أو قواعد البيانات على الويب التي لا تستطيع برامج الزاحف في محركات البحث الوصول إلى محتواها الداخلي، ومن ثم يطلق عليها الويب غير المرئية؛ لكونها لا ترى من جانب محركات البحث. ويطلق عليها أيضًا الويب العميقة والويب الخفية.

٧- **الويب المرئية** : مجموعة صفحات ومواقع الويب التي يصل إليها برنامج الزاحف ويستطيع نسخها إلى مستودع المعلومات في محرك البحث. ويطلق عليها أيضًا الويب السطحية.

٨- **الوثائق** : مجموعات المعلومات أو مصادر المعلومات التي يقتنيها نظام الاسترجاع، وقد تكون الوثائق في شكل تسجيلة أو تسجيلة ومستخلص أو نصوص كاملة.

٩- **التفاعلية** : هي أهم الملامح التي تتميز بها نظم استرجاع معلومات الويب، حيث التحاور والتفاعل مع المستفيد في تلقي الاستفسارات البحثية وتقديم النتائج وتهذيب هذه النتائج مرة أخرى لتحقيق متطلبات المستفيدين من المعلومات.

الديناميكية : ثاني العناصر المهمة في نظم الويب، وتتمثل في سرعة تغيير نظم الويب لمحتواها وتغيير أشكال الوثائق عند استرجاعها، إلى جانب الحركة الداخلية التي تموج بها الويب في تغيير وتحول المحتوى.

الفصل الأول

التعريف بنظم استرجاع المعلومات

التعريف بنظم استرجاع المعلومات :

١ / ٠ التمهيد:

ترتبط نظم استرجاع المعلومات بظهور المعلومات المسجلة ؛ حيث يستمد نظام استرجاع المعلومات مفهومه من عمليات المعالجة والتنظيم التي تتم لمجموعة من المعلومات المسجلة. ومن ثم فإن مكتبة الإسكندرية القديمة هي نظام لاسترجاع المعلومات المسجلة على اللفافات الورقية وقتئذ. ويختص هذا الفصل باستعراض الجانب التاريخي لتطور نظم استرجاع المعلومات، مع التعرض للمكونات والعناصر التي يبنى عليها نظام استرجاع المعلومات كمختلف النظم مثل؛ المدخلات والمخرجات والعمليات الفنية. وقد طرأت على نظم استرجاع المعلومات ومكوناتها الكثير من التغييرات سواء أكانت في المفاهيم أم في أداء العناصر الفرعية وسير العمليات داخل النظام، ومن ثم يستعرض هذا الفصل هذه التغييرات، ملقيًا الضوء على الدور الذي أدته الويب بوصفه نظامًا لاسترجاع المعلومات الرقمية في تغيير أنماط وأشكال أداء نظم استرجاع المعلومات الأخرى.

١ / ١ ماهية نظم استرجاع المعلومات:

١ / ١ / ١ المفهوم والدلالة:

إن مصطلح " نظم استرجاع المعلومات" يحوي الكثير من الدلالات الخاصة بمجال المكتبات والمعلومات، غير أنه كباقي المصطلحات الدالة ينتهي إلى مفهوم محدد لتوضيح معنى بذاته. ولتوضيح هذا الأمر فإننا عندما نتأمل الكلمات المكونة للمصطلح نجد أنه يتكون من كلمات "نظم"، "استرجاع"،

"المعلومات". وتشكل كل كلمة من الكلمات السابقة مصطلحًا قائمًا بذاته يحمل دلالة محددة.

- يشير مصطلح النظام إلى مجموعة العناصر الفرعية المتراكبة المتكاملة المتفاعلة معًا لأداء مهام محددة.
- ويشير مصطلح الاسترجاع إلى استدعاء الأشياء التي تم اختزانها مسبقًا.
- ويشير مصطلح المعلومات إلى تلك الظاهرة التي تغير من الحالة المعرفية لدى متلقيها دائمًا بالإيجاب.

وعلى ذلك فإن مصطلح نظم استرجاع المعلومات يأخذ دلالته وفحواه من تكامل معنى ومفهوم هذه المصطلحات الثلاثة. وقد تناولت مجموعة من التعريفات مفهوم نظم استرجاع المعلومات ومن بينها:

تعرف الموسوعة العربية لمصطلحات المكتبات والمعلومات والحاسبات نظم استرجاع المعلومات على أنها " مجموعة من الإجراءات الميكنة عادة تستخدم في الرجوع إلى البيانات التي تحويها الوثائق وتكشيف تلك البيانات واختزانها بطريقة يمكن استعادتها عند الطلب".

كما يشير علي بن سليمان الصوينع إلى أن نظام استرجاع المعلومات "هو النظام الذي يتعامل مع معلومات وسيطة بين المستفيد والمعلومات النهائية التي يسعى إليها، فهو نظام لا يزيد الحاجة المعرفية عند المستفيد، ولكنه يحيط المستفيد علمًا بوجود معلومات عن الوثائق التي تتعلق بموضوع استفساره". غير أن الصوينع في هذا التعريف خصص هذا المفهوم على نظم استرجاع المعلومات

تلك التي تحوي التسجيلات الببليوجرافية أو التسجيلات والمستخلصات، دون أن ينسحب المفهوم ليشمل نظم الاسترجاع التي تحوي النصوص الكاملة لمصادر المعلومات.(١)

وقد أشار لانكستر في حديثه عن نظم استرجاع المعلومات إلى أن هناك فارقًا بين ما يعرف باسترجاع المعلومات ونظم إيصال الوثائق؛ حيث يقتصر دور الأولى على إحاطة المستفيد بإشارات ببليوجرافية عن الوثائق التي تلبي حاجاته الموضوعية دون الوصول إلى الوثائق ذاتها، أما نظم إيصال الوثائق فهي تلك النظم التي تعطي المستفيد خدمة معلوماتية كاملة، حيث توفر له الوصول إلى مصادر المعلومات الأولية(٢).

ويعرف قاموس مصطلحات المكتبات وعلم المعلومات على الخط المباشر ODLIS عملية استرجاع المعلومات Information retrieval على أنها " العملية أو الإجراء أو الطريقة التي تنتقي مجموعة من المعلومات المسجلة لاستدعائها بحسب مجموعة من المصطلحات البحثية التي تعبر عن حاجة موضوعية محددة، وهذه العملية قد تكون تقليدية كما يتم في استخدام الفهارس البطاقية في المكتبات أو قد تكون إلكترونية كما يتم في الفهارس على الخط المباشر OPAC أو مراصد البيانات الببليوجرافية Bibliograghic databases."(٣)

---

(١) الصوينع، علي بن سليمان. **استرجاع المعلومات في اللغة العربية** .- الرياض : مكتبة الملك فهد الوطنية، ١٩٩٤، ص ٦٨-٧٧ .

(٢) لانكستر، ولفرد. **نظم استرجاع المعلومات** ؛ ترجمة حشمت قاسم. - القاهرة : دار غريب، ١٩٨١.

(٣) Reitz, Joan M. ODLIS —Online Dictionary for Library and Information Science, 2006, cited at 1/12/2007, cited at http://lu.com/odlis/odlis_e.cfm# information retrieval

وبالاعتماد على ما سبق فإننا يمكن أن نضع تعريفًا لنظم استرجاع المعلومات على أنها " هي النظم التقليدية أو الرقمية المكونة من مجموعة من العناصر أو البرامج الفرعية التي تعمل على إضافة واختزان وتنظيم ومعالجة المعلومات حتى يتم استرجاعها مرة أخرى وفقًا للحاجة الموضوعية للمستفيدين."

١ / ١ / ٢ عوامل الاحتياج إلى نظم استرجاع المعلومات :

يهدف علم المعلومات والمكتبات دائمًا إلى الاهتمام بالمعرفة المسجلة اقتناءً وتنظيمًا وبحثًا وبثًا ، على اختلاف أشكال ووسائط هذه المعرفة المسجلة. ومن ثم فإنه متى وجدت المعلومات المسجلة كانت الحاجة للتعامل معها وتنظيمها بغرض إعادة بثها مرة أخرى وفقًا لما تتطلبه الحاجة الموضوعية. وقد تطورت الحاجة لنظم استرجاع المعلومات مع تطور وتقدم أساليب الكتابة والنشر والطباعة ثم النشر الإلكتروني. وتحمل كل مرحلة من مراحل تطور النشر احتياجًا أقوى إلى نظم استرجاع المعلومات، فمع زيادة المعلومات المسجلة المتداولة في بيئة النشر تزداد الحاجة دائمًا إلى السيطرة على كم الإنتاج الفكري المنشور عالميًا. وعلى ذلك يمكن صياغة عوامل الاحتياج لنظم استرجاع المعلومات في العوامل التالية:

١- التقدم الهائل في إمكانية النشر الرقمي وإنتاج المعلومات باستخدام البيئة الشبكية العالمية.

٢- التعدد والتنوع في أشكال مصادر المعلومات التي يتم إنتاجها في سوق النشر من دوريات ورسائل وكتب ... الخ.

٣- استحداث أشكال جديدة من مصادر المعلومات سواء ذات الأصل التقليدي مثل الموسوعات على الأقراص المدمجة أو التي استحدثت

تمامًا في البيئة الرقمية كالرسوم الحيوية أو قواعد البيانات في الشكل الرقمي فقط.

٤- السرعة التي طرأت على دورة المعلومات بحثًا وتأليفًا ونشرًا جراء استخدام الحاسبات الآلية، مما جعل المستفيدين في حاجة دائمة للوصول إلى المعلومات الحديثة فور نشرها، ومن ثم كانت الحاجة لنظم تفاعلية تلبي الحاجات الموضوعية الملحة.

وبوجه عام يمكن القول إن نظم استرجاع المعلومات قد وجدت لتؤدي مجموعة من المهام لاستكمال دورة المعلومات بين المؤلف والقارئ، فضلا عن كونها الوسيط بين المستفيد ومصادر المعلومات، وهذه المهام هي:

١- إضافة واقتناء المعلومات في أشكالها المختلفة وفي موضوعاتها المختلفة في حالة الشمول والإحاطة، أو التخصيص في حالة نظم استرجاع المعلومات المتخصصة.

٢- القيام بعمليات التنظيم والمعالجة للمعلومات على اختلاف مستوياتها سواء أكانت تسجيلات ببليوجرافية أو نصوصًا كاملة أو مستخلصات.

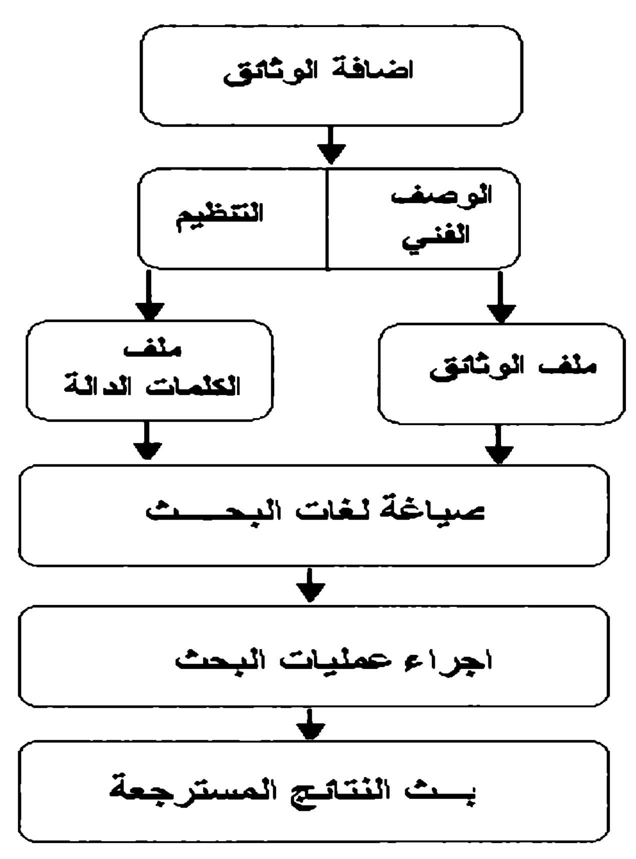
٣- توفير آليات إيصال المعلومات وإتاحتها كالوصول إلى المستفيد من خلال الخط المباشر أو الويب.

٤- توفير الموارد البرمجية أو الموارد البشرية التي تساعد على تحقيق أفضل نتائج للمستفيد عند بحث واسترجاع المعلومات من النظام.

٥- العمل على التطوير والتحديث الدائم سواء أكان ذلك للمعلومات المختزنة أم لاستراتيجيات ولغات البحث.

٦- تقديم المعلومات المسترجعة والنتائج بما يلائم تعامل المستفيد معها والاستفادة منها.

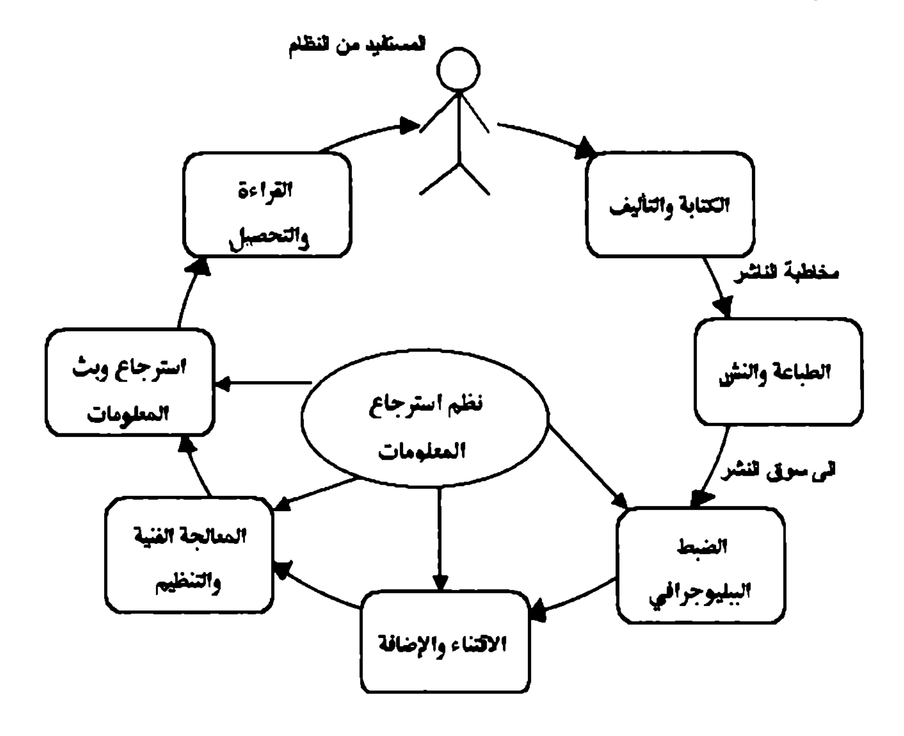
ويمكن تمثيل تدفق العمليات والمهام داخل نظم استرجاع المعلومات على النحو التالي:



الشكل رقم (١ - ١) تدفق العمليات والمهام داخل نظم استرجاع المعلومات

١ / ١ / ٣ نظم الاسترجاع في دورة المعلومات :

تدور المعلومات في سياق دورة محددة متعددة المراحل، تبدأ عند تخليق المعلومات نتيجة لعمليات التأليف والابتكار، ثم تنتهي عند المستفيد من خلال عمليات القراءة والتعرف إلى ما انتهى إليه العلم في المجالات المتخصصة. وكون نظم استرجاع المعلومات تشتغل بمعالجة وتنظيم المعلومات، فإنها تدخل ضمن دورة المعلومات مؤدية مجموعة من المهام تشكل كل منها واحدة من المراحل التي تمر بها المعلومات من البداية إلى النهاية. كما هو معروف إن نظم استرجاع المعلومات تتعامل فقط مع المعلومات المسجلة التي يمكن حصرها وضبطها وتنظيمها ثم بثها في مختلف أشكال مواد ووسائط حمل المعلومات. ومن ثم فإن أولى مراحل ظهور نظم استرجاع المعلومات هي عمليات الإضافة الاقتناء للمعلومات داخل النظام، غير أن بعض نظم المعلومات وخاصة الرقمية تبدأ عملها عند عمليات الضبط الببليوجرافي كونها تمثل نظامًا قائمًا على تنظيم وبث البيانات الببليوجرافية. وتسمى المعلومات عند إضافتها إلى نظام استرجاع المعلومات بالمدخلات ثم تقوم عليها عمليات المعالجة والتنظيم، وتكون مخرجات هذه العمليات هو ما يحصل عليه المستفيد من معلومات في صورة خدمات.



الشكل رقم (١ – ٢) مهام نظام الاسترجاع داخل دورة المعلومات

يوضح الشكل رقم (١ – ٢) مهام نظام استرجاع المعلومات المتمثلة في الآتي:

أولاً : التحكم والضبط بالإنتاج الفكري حديث النشر؛ لتقديم الخدمات الببليوجرافية التي تمكن الباحثين من التعرف إلى المعلومات المنشورة حديثًا، مثل الببليوجرافيات والكشافات والمستخلصات... وغيرها. كما تعد نظم الاسترجاع الببليوجرافية أدوات اختيار وانتقاء لمصادر المعلومات لمرافق المعلومات التقليدية، وهو ما يستفيد منه اختصاصيو وأقسام التزويد.

ثانيًا : إضافة واقتناء المعلومات وهو ما يحدث في مرافق المعلومات بصورة التزويد وتنمية المقتنيات التقليدية، غير أن الويب كنظام استرجاع للمعلومات

الرقمية قد تهتم بعمليات الإضافة على حساب الاقتناء كما يحدث في محرك بحث الويب الذي يعتمد على برنامج الزاحف في عمليات الإضافة لصفحات الويب دون الاهتمام بقيمة أو نوعية المعلومات الحاصل عليها، أما نظم أدلة بحث الويب فتغلب جانب الاقتناء على عملية الإضافة نظرًا لاستخدامها العنصر البشري.

ثالثًا : ممارسة مهام المعالجة الفنية والتنظيم التي تمثل اختلافًا جليًا بين نظم التقليدية والويب ؛ ففي الوقت الذي تقوم فيه حقول الوصف الفني أو البيليوجرافي بتسجيل ملامح وخصائص مصادر المعلومات في النظم التقليدية، تقوم هذه الحقول بأداء دور أكبر في عملية التنظيم الموضوعي ؛ حيث تنشط مختلف هذه الحقول ويمكن البحث داخلها، مما يمكن للمستفيد استخدام كلمات البحث مثل : (ملون، ٢٤٠ صفحة، علم المعلومات وأحمد بدر) جميعا ككلمات مفتاحية لاسترجاع مصدر معلومات محدد.

رابعًا : تقديم خدمات الاسترجاع وبث المعلومات من خلال تخصيص مجموعة من آليات البحث ولغات صياغة استراتيجيات بحث المعلومات، ثم تقديم المعلومات المسترجعة في شكل قوائم من النتائج المسترجعة أو في شكل خدمات معلومات مثل البث الانتقائي والتصوير والاطلاع الداخلي... وغيرها من خدمات نظم المعلومات التقليدية والرقمية.

١ / ٢ تاريخ نظم الاسترجاع :

لقد مرت نظم استرجاع المعلومات بمراحل متعددة من التطور، لم تتوقف فيها على تطور أشكال مصادر المعومات التي تحويها فقط، وإنما امتد هذا

التطور إلى الآليات المستخدمة في إجراء التخزين والاسترجاع للمعلومات. وتعكس المراحل التي ذكرها لانكستر ملامح التطور التي طرأت على نظم الاسترجاع وصولاً إلى أدوات بحث الويب.[١]

- **ما قبل الأربعينات : الاعتماد على نظم الربط المسبق.**

اعتمدت نظم التخزين والاسترجاع للمعلومات على التعامل مع البيانات البليوجرافية تنظيمًا وتخزينًا واسترجاعًا. وقد استخدمت أدوات الضبط والتقنين من قوائم رؤوس الموضوعات في صياغة المداخل الموضوعية المكونة من كلمة واحدة أو مجموعة من الكلمات تنظم جميعها بحسب صياغتها وترتيبها في التعبير عن الموضوع. وفي هذه الحالة، فإنه على المستفيد متابعة سياق الكلمات وتتابعها للوصول إلى الكلمة أو المصطلح الدال على الموضوع.

فمثالاً : يتكون مفهوم " نظم استرجاع المعلومات" من ثلاث كلمات مرتبة ترتيبًا محددًا كي تعكس المعنى الدلالي للمصطلح. ومن ثم لا يمكن الوصول إلى كلمة المعلومات دون المرور هجائيا على كلمتي نظم و استرجاع. ولذلك كان على المستفيد الالتزام بالربط السابق بين الثلاث كلمات وتتابع الوصول إليها بحسب ترتيبها في السياق والترتيب الهجائي.

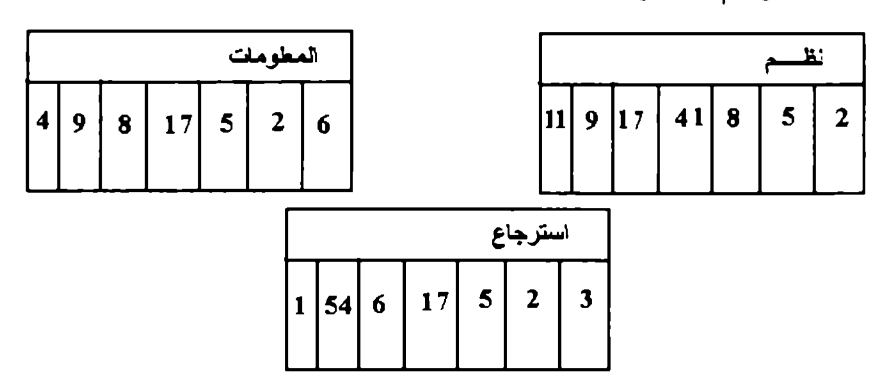
- **مرحلة الأربعينات : مرحلة كشافات الربط اللاحق، استخدام المصطلح الموحد.**

اعتمدت هذه المرحلة على استخدام ما يسمى بالمصطلح الموحد Uni Term. ويعني ذلك أن يكون لكل مصطلح أو رأس موضوع بطاقة خاصة به تجمع

---

(١) لانكستر، ولفرد. مصدر سابق. ص ٤٠.

بداخلها أرقام كل الوحدات أو الوثائق التي تنتمي إلى هذا الموضوع. وعلى ذلك تغير الأمر من بطاقة لكل كتاب إلى بطاقة لكل موضوع. ويتم الربط هنا من خلال المستفيد ذاته، فما عليه في هذه الحالة سوى تحديد المصطلح أو المصطلحات المعبرة عن موضوعه ثم تجميع بطاقات الموضوعات ومضاهاتها لتحديد الأرقام المشتركة فيما بينها.



استرجاع المعلومات باستخدام بطاقات المصطلح الموحد

الشكل رقم (١ – ٣) استرجاع المعلومات باستخدام المصطلح الموحد

ويوضح الشكل رقم (١ – ٣) التوافق بين الكلمات البحثية التي استخدمها المستفيد في الوثائق أرقام : ٢ ، ٥ ، ١٧ أي أن هذه الوحدات هي التي تتحدث عن نظم استرجاع المعلومات. فما على المستفيد هنا سوى تحديد المصطلحات البحثية ومتابعة التوافق في الأرقام التي تدل على الوثائق التي تلبي حاجته الموضوعية.

- **مرحلة الخمسينات : مرحلة البطاقات المثقوبة.**

تستخدم البطاقات المثقبة إلى الآن في الانتخابات الأمريكية، حيث تعتمد فكرتها على مجموعة من الثقوب التي تسمح بنفاذ الضوء فقط بين البطاقات

المتوافقة في الثقوب، مما يدل على التوافق بين المصطلحات التي تخدم الموضوع الواحد والدلالة على الوثائق الملائمة للاسترجاع.

- **مرحلة الستينات : النظم الآلية المعتمدة على الأشرطة.**

جاءت الأشرطة الممغنطة في الستينات لمعالجة العيوب الناتجة عن الأشكال السابقة من نظم الاسترجاع. وقد استخدمت هذه الأشرطة لتقديم بعض خدمات المعلومات المستحدثة مثل إنتاج الكشافات، وخدمة البث الانتقائي للمعلومات SDI ، وتقديم خدمة البحث الراجع باتباع نظام البحث على دفعات batch system.

- **مرحلة السبعينات : النظم الآلية على الخط المباشر.**

عملت الحاسبات الآلية على تقدم نظم الاسترجاع من خلال الاعتماد على إنتاج الأقراص المضغوطة أو استخدام البحث على الخط المباشر والاعتماد على شبكات الربط بين الفهارس الآلية. وهو ما أتاح الاسترجاع في وقت إجراء البحث خلاف ما حدث في البحث على دفعات. وفي هذه المرحلة خرجت آليات ومعاملات البحث لتساعد على تقدم صياغة استراتيجيات البحث.

- **مرحلة التسعينات : اللاورقيات.**

ساعد التطور في تقنيات الحاسبات الآلية والقدرات الفائقة على التخزين والاسترجاع أن يبدأ التحول من الشكل الورقي المطبوع إلى النص الكامل الرقمي. بالإضافة إلى آليات بحث النصوص الكاملة، ومن ثم يستطيع المستفيدون الآن استرجاع الوثائق أو النصوص المختلفة بأي من الكلمات الواردة في هذه النصوص من خلال وضع كلمات النص جميعها كنقاط وصول إلى هذه الوثائق.

- **المرحلة الحالية : آليات الاسترجاع على الويب.**

يمكن القول الآن أن الويب كإحدى خدمات الإنترنت تشكل نظامًا كونيا لاسترجاع المعلومات؛ حيث يتم حفظ واختزان المعلومات بكافة أشكالها ووسائط حملها داخل مستودعات خادمات الويب web servers. إلى جانب اعتمادها لآلية التنظيم باستخدام بنية النص الفائق hyper text المعتمد على روابط links تصل بين مجموعة من العقد nodes. ومن ثم يستطيع المستفيد أو مستخدم الويب استرجاع المعلومات في فضاء الويب باستخدام أسلوب التصفح كأحد أساليب الاسترجاع في بيئة الويب. إلى جانب ذلك فإن الويب تحمل بداخلها نظمًا لاسترجاع المعلومات تتجه إلى التخصص في مجالاتها الموضوعية، أو في وسائط حمل المعلومات كنظم استرجاع النصوص والصوت والصورة الثابتة أو المتحركة، أو في بنيتها وتقنيات التنظيم والاسترجاع كمحركات وأدلة وبوابات البحث.

ويمكن القول نهاية أن تطور نظم استرجاع المعلومات ساعد على تحقيق الجوانب التالية:

أولاً : إمكانية وصول المستفيدين إلى نظم الاسترجاع عن بعد من خلال الشبكة العالمية وإجراء عمليات البحث والاسترجاع للمعلومات التي تلبي الحاجات الموضوعية.

ثانيًا : إمكانية الربط بين التسجيلات الببليوجرافية التي تعطي مجرد إشارات عن المعلومات الأولية، بالنصوص الكاملة لها ومن ثم يستطع المستفيد الوصول إلى المعلومات الأولية التي يحتاجها دون التوقف عند الإشارات الببليوجرافية.

ثالثًا : عملت الويب على استحداث مجموعة من نظم استرجاع المعلومات متمثلة في أشكال ووسائط حمل معلومات حديثة لم تكن توجد في البيئة التقليدية أو حتى الإلكترونية تستطيع التعامل معها ، وهي مثل محركات بحث الصوت والصورة كنظم لاسترجاع الصوت والصورة، هذا إلى جانب الحديث عن نظم استرجاع قواعد البيانات ذاتها وإمكانية الجمع بين قواعد بيانات ببنية مختلفة واسترجاع المعلومات منها في وقت واحد.

١ / ٣ أنواع نظم استرجاع المعلومات:

لقد حصرت الكتابات التي تحدثت عن أنواع نظم استرجاع المعلومات في ثلاثة أنواع رئيسة تستمد تقسيمها من طبيعة البيانات أو المعلومات التي تحويها هذه النظم وتتمثل هذه الأنواع في[(١)] :

١- نظم استرجاع المعلومات :

وهي النظم التي تستخدم لاسترجاع الإشارات الببليوجرافية المنفردة أو تلك المصاحبة للمستخلصات.

٢- نظم استرجاع الحقائق :

وهي النظم التي تسترجع المعلومات المختصرة مثل الإحصاءات ومقالات الموسوعات وكتب التراجم.

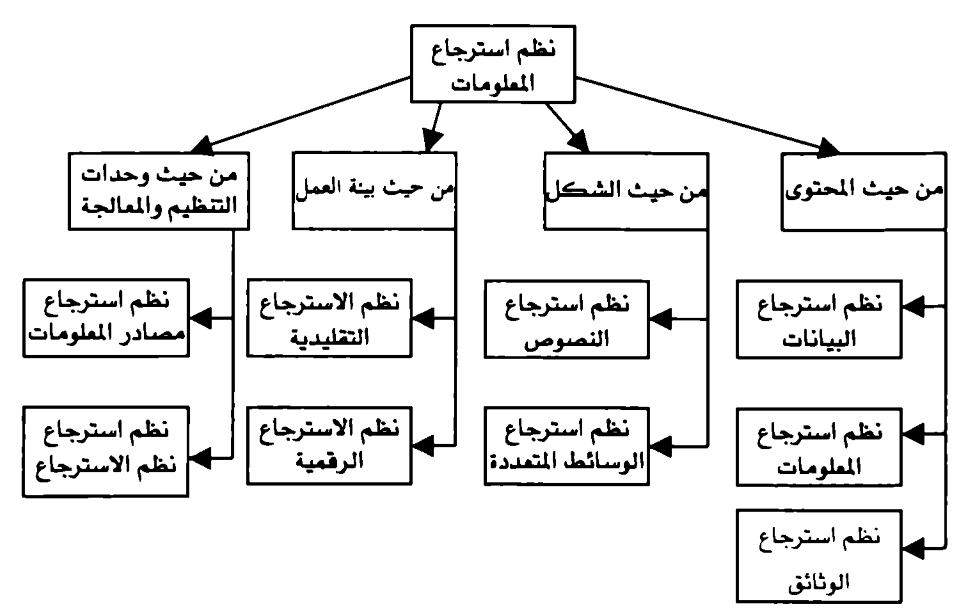
---

(١) فاتن سعيد بامفلح. **أساسيات نظم استرجاع المعلومات الإلكترونية** .- الرياض : مكتبة الملك فهد الوطنية، ٢٠٠٧. متاح على
http://www.kfnl.gov.sa/idarat/alnsherel/asaseiatnzmstrgaaalmalomat/pdf/f1.pdf

٣- نظم استرجاع الوثائق :

وهي النظم التي تسعى للجمع بين الإشارات الببليوجرافية والنص الكامل لمصدر المعلومات، حتى يتمكن المستفيد من الحصول على المحتوى في نفس وقت حصوله على البيانات الببليوجرافية.

غير أن الباحث يرى أن هذا التقسيم إنما يدل على منظور ضيق ينظر من خلاله لنظم استرجاع المعلومات ؛ حيث إن نظم استرجاع المعلومات قد اكتسبت الكثير من الملامح والخصائص التي طرأت علها جراء التطور في الحاسبات الآلية وشبكات المعلومات. ومن ثم فإننا نحاول جاهدين أن ننظر لنظم استرجاع المعلومات بمزيد من الاتساع، بما يمكننا من تغطية مختلف جوانبها ورصد أنواعها المتعددة. غير أن الكاتب سيستخدم مصطلح نظم استرجاع المعلومات في سياق هذه الدراسة، لما لهذا المصطلح من شمولية. وتتمثل وجهة النظر هذه في تبني التقسيم التالي:



الشكل رقم (١ – ٤) أنواع نظم استرجاع المعلومات

يوضح الشكل رقم (١ – ٤) التقسيم العام لنظم استرجاع المعلومات، وقد اهتم الكاتب بذكر الأنواع الأخرى من نظم الاسترجاع دون التقسيم الموضوعي لنظم الاسترجاع إلى متخصصة وشاملة ؛ وذلك لكونه تقسيمًا بديهيًا لا يحتاج التطرق إليه في الحديث. وجاءت باقي الأنواع كالتالي:

١ / ٣ / ١ من حيث المحتوى :

تعتمد نظم استرجاع المعلومات على مستودع أو قاعدة بيانات لاختزان وحدات المعلومات، وهذه الوحدات تختلف في درجة إفادتها لمستخدمي نظم الاسترجاع، حيث توجد الإشارات الببليوجرافية التي تقدم فقط بيانات أو حصرًا بالوثائق التي تلبي الحاجة الموضوعية، أو تزيد عن ذلك بوجود مستخلص للوثيقة المصدرية، أو أن تحتوي الوثائق الصدرية ذاتها. ومن ثم انقسمت إلى:

١ / ٣ / ١ / ١ نظم استرجاع البيانات:

هي تلك النظم التي تختزن الإشارات الببليوجرافية في شكل تسجيلات مكونة من مجموعة من الحقول، ويمثلها فهارس المكتبات على الخط المباشر Online Public Access Catalog OPAC ، أيضا قواعد البيانات الببليوجرافية على الأقراص المضغوطة CDs مثل قاعدة البيانات books in print، إلى جانب نظام استرجاع البيانات الببليوجرافية الشبكي OCLC الذي يهدف إلى توحيد وتبادل التسجيلات الببليوجرافية بين المكتبات.

- الفهارس المحسبة على الخط المباشر OPAC.
- نظم تبادل التسجيلات OCLC.

١ / ٣ / ١ / ٢ نظم استرجاع المعلومات:

تختلف نظم استرجاع المعلومات عن سابقتها في تقديم المعلومات الأولية التي تلبي الحاجة الموضوعية مباشرة، غير أن المعلومات الأولية التي تحويها هذه النظم تختلف عن نظم النصوص الكاملة والوثائق في أنها تشبه معلومات الخدمة المرجعية، حيث تتسم بالإيجاز والتلخيص والسرعة مثل مقالات الموسوعات وكتب التراجم وبيانات الإحصاءات. ويمكن القول إن الاختلاف فقط في شكل المعلومات المقدمة. مثل: -

- مواقع الموسوعات وكتب التراجم مثل biography.

١ / ٣ / ١ / ٣ نظم استرجاع الوثائق "النصوص الكاملة":

تتميز نظم استرجاع الوثائق أو النصوص الكاملة بتقديم ميزة البحث في النص الكامل واسترجاع الوثائق باستخدام مختلف الكلمات الواردة في المتن؛ بالإضافة إلى أنها تلغي الدور الوسيط للإشارات الببليوجرافية بين المعلومات الأولية وبين المستفيد، حيث يمكن للمستفيد الحصول على حصر بالوثائق التي تفيد موضوع البحث إلى جانب الحصول على متن ونصوص هذه الوثائق في الوقت نفسه. مثل:

- قواعد بيانات النصوص الكاملة مثل شركة wilson.

١ / ٣ / ٢ من حيث الشكل :

منح الحاسب الآلي نظم الاسترجاع القدرة على اختزان مختلف أشكال المعلومات، حتى غير المستخدم منها في البيئة التقليدية مثل المواد المصورة والمواد الصوتية. إلى جانب استحداثه أشكالاً أخرى مثل الرسوم الحيوية Animations.

١ / ٣ / ٢ / ١ نظم استرجاع النصوص :

تتعامل هذه النظم مع المواد النصية، ومن ثم تكون الكلمة هي الوحدة المختزنة وأداة العمل لاسترجاع هذه الوحدات مرة أخرى؛ حيث تستخدم استراتيجيات البحث بالكلمات المفتاحية لاسترجاع هذه النصوص. مثل:

- محركات بحث صفحات الويب.

- قواعد بيانات النصوص الكاملة مثل شركة wilson.

١ / ٣ / ٢ / ٢ نظم استرجاع الوسائط المتعددة:

تشمل هذه النظم أنواعًا مختلفة يصعب حصرها من نظم الاسترجاع التي تتعامل مع الوسائط المتعددة، وهذا إنما يأتي من التنوع الكبير والتعددية التي توجد عليها الوسائط المتعددة، حيث تشمل الصوت والصورة الثابتة والمتحركة ومواد الرسوم الحيوية ... وغيرها. مثل:

- محركات بحث الصور مثل Picsearch.[١]

- محركات بحث الصوت مثل all music.[٢]

١ / ٣ / ٣ من حيث بيئة العمل:

إنه من الخطأ أن نقصر وجود نظم استرجاع المعلومات على وجود الحاسب الآلي، فإن نظم الاسترجاع ليست فقط هي النظم الإلكترونية أو الرقمية في اختزان وتنظيم واسترجاع وحدات المعلومات. وإنما تشمل نظم الاسترجاع أيضا النظم التقليدية والمطبوعة.

(١) www.picseach.com

(٢) www.allmusic.com

١ / ٣ / ٣ / ١ نظم الاسترجاع التقليدية :

تحوي هذه النظم تسجيلاتها ووحداتها في شكل مطبوع أو ورقي، وأهم هذه النظم الفهارس البطاقية في المكتبات والببليوجرافيات الورقية المطبوعة على اختلاف أنواعها. مثل:

- الفهارس البطاقية في المكتبات.
- الببليوجرافيات الورقية.

١ / ٣ / ٣ / ٢ نظم الاسترجاع الرقمية :

ينسحب مفهوم نظم الاسترجاع الرقمية على مختلف نظم استرجاع المعلومات التي تعتمد على العمل في بيئة تقليدية باستخدام الحاسب الآلي مثل الفهارس الآلية للمكتبات أو تلك التي توجد في بيئة رقمية ولا يوجد لها واقع مادي مثل محركات البحث على الإنترنت أو قواعد بيانات الإنترنت. مثل: -

- محركات البحث على الويب.
- الفهارس الآلية على الخط المباشر OPAC.
- قواعد بيانات الأقراص المضغوطة.

١ / ٣ / ٤ من حيث وحدات التنظيم والمعالجة:

تختلف نظم استرجاع المعلومات فيما بينها من حيث الوحدات المختزنة والمعالجة داخل نظم الاسترجاع، فمن الممكن أن تتمثل هذه الوحدات في شكل مصدر معلومات أولي من النصوص الكاملة أو ثانوي في شكل تسجيلات ببليوجرافية. غير أن الحاسبات الآلية قدمت لنا نوعًا جديدًا من نظم الاسترجاع يتمثل في نظم استرجاع نظم الاسترجاع.

١ / ٣ / ٤ / ١ نظم استرجاع مصادر المعلومات:

يتمثل هذا النوع في نظم الاسترجاع في طبيعتها الحالية التي تحوي مجموعة المعلومات المنظمة المختزنة؛ ليتم استرجاعها مرة أخرى مهما اختلفت هذه المعلومات بين الإشارات الببليوجرافية أو النصوص أو الوسائط المتعددة. مثل:

- فهارس المكتبات.
- قواعد بيانات النصوص الكاملة مثل wilson databases.

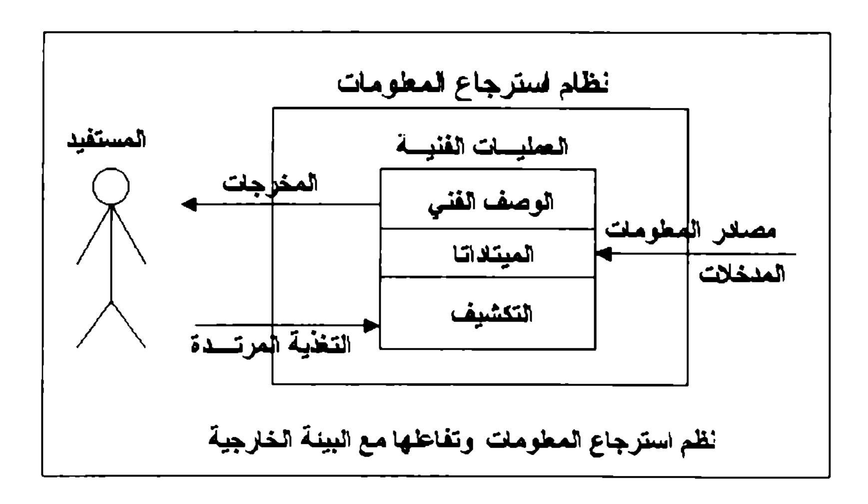
١ / ٣ / ٤ / ٢ نظم استرجاع نظم الاسترجاع:

لا يحتوي هذا النوع من نظم الاسترجاع على وحدات من المعلومات أو تسجيلات ببليوجرافية، وإنما يحتوي على نظم استرجاع كاملة ؛ حيث يعمل النظام الأكبر والرئيس على اختزان أو تجميع مجموعة من نظم الاسترجاع ثم العمل على تكشيفها وتنظيمها ثم استرجاع محتوياتها ومن خلال واجهة أو نافذة واحدة لبحث مختلف نظم الاسترجاع الداخلية في وقت واحد. ومجازًا يمكن تشبيه نظم استرجاع نظم الاسترجاع بمثال ببليوجرافيات الببليوجرافيات، غير أن هذا النوع من النظم يستطيع استرجاع محتويات النظم الداخلية فيه بعكس ببليوجرافيات الببليوجرافيات التي تحتوي فقط على بطاقات الوصف الفني عن الببليوجرافيات التي تضمها. وأفضل أمثلة هذا النوع:

- محركات البحث المتعددة meta search engines.
- محركات بحث قواعد البيانات databases search engines.

١ / ٤ مكونات نظم استرجاع المعلومات:

يتكون نظام استرجاع المعلومات كغيره من النظم من مجموعة من المكونات التي تتفاعل ولا يمكن لهذا النظام العمل بدون أي من هذه المكونات. وكما هو معلوم، فإن مكونات النظم تختلف فيما بينها بحسب طبيعة البيئة التي يعمل فيها النظام ؛ فهناك النظم التي تعمل في بيئة مغلقة بما لا تحتوي معها مدخلات أو مخرجات وتغذية مرتدة. وهناك النظم التي تعمل في بيئة مفتوحة تأخذ منها المدخلات الواردة إلى النظام وتدفع إليها المخرجات الناتجة عن العمليات الداخلية للنظام. وتندرج نظم استرجاع المعلومات أسفل نوعية النظم المفتوحة؛ لكونها تستمد مدخلاتها الممثلة في مصادر المعلومات من سوق النشر التقليدي أو الإلكتروني، وتعطي المخرجات الممثلة في معلومات أولية أو إشارات ببليوجرافية منظمة. ويوضح الشكل رقم (١ – ٥) مكونات نظام استرجاع المعلومات في ضوء علاقته بالبيئة الخارجية:



الشكل رقم (١ – ٥) العلاقة بين نظم استرجاع المعلومات والبيئة الخارجية

يوضح الشكل رقم (١ — ٥) مكونات نظم استرجاع المعلومات المتمثلة في التالي:

١ / ٤ / ١ المدخلات :

يستقبل نظام الاسترجاع مجموعة متنوعة من المدخلات تختلف في أهميتها لنظام المعلومات، غير أن الأهمية القصوى تعود إلى مصادر المعلومات التي تستقبلها نظم استرجاع المعلومات من سوق النشر المحلي أو العالمي. وكما تختلف مصادر المعلومات في الشكل والنوع التقليدي أو الرقمي، فإنها تختلف أيضًا في طبيعة إدخالها إلى النظام، فقد تكون الوسيلة هي أوامر التوريد كما في المكتبات التقليدية وقد تكون خدمات الفهرسة الآلية OCLC هي المزود بهذه المدخلات، وهناك من النظم ما يبنى عمله على السعي إلى مصادر المعلومات ذاتها وانتقائها ثم معالجتها كما يفعل محرك البحث في بناء برامج الزاحف في الوصول إلى صفحات الويب وضمها إلى مستودع محرك البحث.

١ / ٤ / ٢ عمليات المعالجة:

تمثل عمليات المعالجة الفنية واسطة العقد في عمل نظام استرجاع المعلومات، ذلك لكون نظام الاسترجاع بلا تنظيم مجرد مستودع لاختزان المعلومات دون التفكير في تيسير عمليات استرجاعها مرة أخرى. وإذا ما حاولنا التعبير عن هذه المعالجة والتنظيم في نظام الاسترجاع فإنها تمثل ما يطلق عليه عمليات القيمة المضافة داخل نظم المعلومات. وقد أخذت المعالجة الفنية والتنظيم مراحل مختلفة من التطور؛ لتصل إلى ما هي عليه الآن في نظم استرجاع معلومات الويب. حيث بدأت المعالجة الفنية والتنظيم لمصادر المعلومات التقليدية اعتمادًا على استخدام الفهرسة الوصفية والموضوعية واعتمادًا على العنصر البشري في الأداء والتجهيز، ثم تطور إلى استخدام الحاسب الآلي في

عمليات الفهرسة الوصفية والموضوعية لتحقيق عوامل السرعة والدقة والتوفير في الحيز، غير أنها مازالت تعتمد في هذه المرحلة أيضًا على العنصر البشري. وقد استقرت نظم استرجاع المعلومات الآن عند تلك المرحلة التي يقوم فيها النظام ذاتيًا بإجراء عمليات المعالجة والتنظيم الموضوعي آليًا باستخدام برامج التكشيف الآلي والاعتماد على عناصر الميتاداتا Metadata أو البيانات الخلفية لمصادر المعلومات الرقمية. وتعد محركات بحث الويب من أفضل نظم الاسترجاع التي توفر المعالجة والتنظيم باستخدام برامج التكشيف الآلي. وتنقسم عمليات المعالجة الفنية إلى شقين هما:

١ / ٤ / ٢ / ١ الوصف الفني:

تطورت عمليات الوصف الفني بداية من الاعتماد على الفهارس البطاقية إلى خدمات تبادل التسجيلات الببليوجرافية على الخط المباشر، ثم الاعتماد على معايير وصف ملفات مصادر المعلومات على الويب باستخدام حقول وعناصر الميتاداتا metadata. لقد اعتمدت المعالجة الفنية لمصادر المعلومات زمنًا طويلاً على استخدام قواعد الفهرسة الأنجلوأمريكية لفهرسة ووصف مختلف مصادر المعلومات التقليدية في مرافق المكتبات. وقد تعددت وتباينت أنواع الفهارس شكلاً وموضوعًا؛ حيث ظهر الفهرس البطاقي والمحزوم والكتاب، كما ظهرت فهارس المؤلف والعنوان الموضوعي والقاموسي والمصنف. وتقدمت باستخدام الحاسب الآلي لتدخل ضمن ما يعرف بالفهرس المتاح على الخط المباشر OPAC. وسوف يقتصر الحديث هنا على بيانات الوصف الخلفية metadata التي تمثل واقع معالجة مصادر المعلومات على الويب، والتعرض لها من خلال التعرف إلى بعض معاييرها.

أولاً : الميتاداتا : metadata

يطلق على الميتاداتا مصطلحات البيانات الخلفية، وما وراء البيانات ، والبيانات عن البيانات، ولإيضاح المعنى أكثر فهي بيانات تصف ملامح وخصائص مصادر المعلومات الرقمية، وتوضح علاقاتها، وتساعد على معالجتها وتنظيمها واسترجاعها مرة أخرى، أي أنها توفر إدارة مصادر المعلومات الرقمية واستخدامها بفعالية. وبعيدًا عن الخوض الكثير في تفاصيل التعريف بالميتاداتا فإنه يمكن القول بأنها " مجموعة من عناصر أو حقول البيانات التي تصف بدقة مختلف ملامح مصادر المعلومات سواء المتعلقة بجوانب الوصف الفني لمصدر المعلومات أو أصل ونشأة ومنتج المصدر أو المتعلقة باستخدام وإدارة حقول ملكية المصدر." وعلى ذلك فإن هذه الحقول تنقسم إلى ثلاثة أنواع أساسية هي: الميتاداتا الوصفية، و الميتاداتا الإدارية، و الميتاداتا البنائية. يمكن تناولها كالتالي[(١)]:

١- الميتاداتا الوصفية: Descriptive Metadata

وهي مجموعة حقول الميتاداتا التي تصف المصادر الرقمية من حيث خصائصها الشكلية ومحتوياتها الفكرية، وهي تتشابه في ذلك مع حقول الوصف الببليوجرافي في بطاقات مثل الفهارس والكشافات، وتضم عناصر مختلفة تساعد على إيجاد المصادر مثل منشئ العمل، والعنوان، والمصطلحات الموضوعية، والتصنيف، والمصادر ذات الصلة ... وغيرها.

---

(١) METADATA. North Carolina university, 2005, cited 20/10/2007, cited at http://www.ncecho.org/Guide/metadata.htm

٢- الميتاداتا الإدارية: Administrative Metadata

وهي مجموعة حقول الميتاداتا التي تستخدم لإدارة المواد وحفظها في المستودع. وتضم الميتاداتا الإدارية المعلومات المتعلقة بإدارة حق المؤلف، والتراخيص والشروط والقيود المرتبطة بهذا الجانب. كما تضم أيضاً المعلومات الفنية الخاصة بنوع الملف، ودرجة الوضوح للصور، والبرامج ، والعتاد المستخدم لإنتاج الصور والألوان؛ أي أن هذه البيانات تختص بالوصف التقني الذي يوجد عليه مصدر المعلومات.

٣- الميتاداتا البنائية: Structural Metadata

وهي مجموعة حقول الميتاداتا التي تستخدم أساساً لتخزين المواد الإلكترونية في المستودع، حيث تساهم في تركيب المحتوى الداخلي لوحدات المصادر الرقمية المكونة لنظام الاسترجاع. كما تساعد على عرض وتصفح المصادر الإلكترونية، كما تسمح بربط كل مادة من المواد بالأخرى لتكون وحدة منطقية من تلك المواد ؛ فهي تمثل معلومات عن البناء والتنظيم الداخلي للمواد الرقمية وترابطها داخل نظام الاسترجاع.

وقد أضافت بعض الكتابات نوعين آخرين من أنواع بيانات وعناصر الميتاداتا هما:

٤- ميتاداتا الحفظ والتسجيل: Preservation metadata

وهي مجموعة حقول الميتاداتا الخاصة بكيفية تخليق وإنتاج وحفظ بعض مصادر المعلومات الرقمية مثل الصور الرقمية الثابتة. فمن بين هذه الحقول بيانات المسح الضوئي وأنواع الملفات الرقمية وأبعاد الشكل المصور وكثافة العرض للشكل المصور.

٥- ميتاداتا تحليل المحتوى: Analytical metadata

وهي مجموعة حقول الميتاداتا الخاصة بمحتوى وموضوع مصادر المعلومات الرقمية ؛ حيث تصف هذه العناصر الموضوع والسياق الوارد في المصدر الموصوف.

وبوجه عام يمكن القول إن مجموعات عناصر الميتاداتا على اختلاف أنواعها تهدف جميعًا إلى تحقيق الأهداف التالية[١]:

١- إعطاء وصف عام شامل للسياق الذي ورد به مصدر المعلومات الرقمي.

٢- توثيق المعلومات الرقمية المتاحة على الويب التي يتم تنظيمها فقط داخل نظم الاسترجاع.

٣- التعريف بالملامح البنائية للعلاقات التي توجد داخل وبين موضوعات مصادر المعلومات الرقمية.

٤- تجهيز مجموعة متنوعة من نقاط الإتاحة والوصول إلى مصادر المعلومات الرقمية أمام المستفيدين.

٥- تحقيق التوافق بين مصادر المعلومات التي تم إنتاجها ونشرها وبين خصائص عمليات التنظيم والاسترجاع في نظم المعلومات في البيئة الرقمية.

---

(١) أمل حسين عبد القادر. **المتطلبات الوظيفية للتسجيلة الببليوجرافية** .- cybrarians journal .- ع ٥ (يونيو ٢٠٠٥) .- تاريخ الاطلاع < اكتب هنا تاريخ اطلاعك على المقالة > .- متاح في: http://www.cybrarians.info/journal/no5/frbr.htm

٦- تجهيز مجموعة من عناصر الوصف المختلفة والمتعددة التي تحقق لاختصاصيي المعلومات السيطرة على كل مصادر المعلومات المنتجة كمًا وكيفًا على تعدد موضوعاتها وأشكالها.

ولعل أهم ما يفيد هنا في الوصف الفني هي مجموعة البيانات الوصفية التي تتعامل مع مصادر المعلومات الرقمية مكونة التسجيلات الببليوجرافية الواصفة والداعمة لتنظيم واسترجاع المصادر من داخل نظام الاسترجاع. أما الأنواع الأخرى، فإنها لا تهم المستفيد كثيرًا، حيث تدخل ضمن بناء وتركيب نظام استرجاع المعلومات.

ثانياً: أهم معايير الميتاداتا:

١) معيار Dublin Core

تم الاتفاق على إصدار هذا المعيار في عام ١٩٩٥م في مدينة دبلن Dublin في ولاية أوهايو Ohio بين مجموعة من الهيئات المنتشرة في مختلف الدول مثل إنجلترا وأستراليا وكندا واليابان وألمانيا وغيرها من الدول. ولقد استخدمت الهيئات غير الرسمية القائمة على تطوير وتحديث بيئة الويب المعايير المختلفة الداعمة لحقول مثل دبلن كور والتي تم تطويرها في اللغة المعيارية الجديدة XML. ويمكن تحديد عناصر الوصف كما جاءت في معيار دبلن كور على النحو التالي(١):

(١) سيد ربيع سيد. **محركات بحث الصور الثابتة على الإنترنت: دراسة تحليلية**.- الرياض: مكتبة الملك فهد الوطنية، ٢٠٠٧.

الجدول رقم (١ –١) حقول الميتاداتا لوصف الملفات الرقمية

DC. Title	عنوان مصدر المعلومات
DC. Creator	المنشئ أو المؤلف لمصدر المعلومات
DC. Subject	الموضوع أو الكلمة الدالة على المحتوى
DC. Description	ملخص أو تعليق على الموضوع
DC.Publisher	الناشر سواء كان شخصًا أو هيئة
DC. Contributor	المساعد في إنشاء المصدر سواء كان هيئة أو شخصًا
DC. Date	تاريخ العمل ويفصل لعدة حقول فرعية أخرى
DC. Type	نوع مصدر المعلومات
DC. Format	Format الشكل المادي أو شكل البنية الرقمية
DC. Identifier	محدد مسار المصدر.URL, URN, ISBN, etc :
DC. Source	المصدر الأساسي المادي الذي تم الاعتماد عليه في إنشاء المصدر الرقمي
DC. Language	لغة مصدر المعلومات
DC. Relation	العلاقات الببليوجرافية مع باقي مصادر المعلومات
DC. Coverage	التغطية المكانية أو الجغرافية للمصدر
DC. Rights Copyright	ملخص لحقوق الملكية الفكرية والمادية لمصدر المعلومات

وقد اعتمد الباحث معيار دبلن كور كمعيار ملائم لاسترجاع الصور الرقمية في عمل محركات أو نظم استرجاع الصور على الويب في دراسته التي تناولت محركات بحث الصور الثابتة على الإنترنت. ويأتي ذلك من الشمولية والتغطية التي يحققها معيار دبلن كور في وصف ملفات الصور الرقمية، حيث يتسم بالملامح التالية:

❖ بلغ معيار دبلن من الانتشار الدرجة التي جعلته معيارًا عالميًا في وصف مصادر المعلومات الرقمية على الويب عامة، والصور الرقمية خاصة.

❖ يشتمل معيار دبلن على مجموعات من عناصر الوصف ذات اتجاهات مختلفة؛ حيث يوفر العناصر التي تصف الشكل المادي للصور كما يوفر العناصر التي تصف المحتوى والموضوع، إلى جانب العناصر الإدارية والتنظيمية لملفات الصور ومنها حقوق الملكية.

❖ إمكانية إصباغ مختلف الملفات الرقمية بعناصر الوصف في دبلن كور حتى تستطيع محركات ونظم الاسترجاع القيام بالأداء النمطي والمعياري في معالجة وتنظيم ملفات الويب الرقمية.

٢- معيار: MARC

يعرف معيار مارك MARC بأنه شكل من أشكال التسجيلات الببليوجرافية التي يمكن قراءتها وتبادلها آلياً باستخدام الحاسبات الآلية، وقد جاء الاسم الاستهلالي من مصطلح Machine Cataloging Readable أي الفهرسة المقروءة آلياً، ويقابل هذا المصطلح مصطلح آخر في اللغة العربية هو (فما) أي الفهرسة المقروءة آليًا. ومعيار مارك عبارة عن التسجيلات الببليوجرافية لمصادر المعلومات التي توجد في الشكل الرقمي خلاف الأشكال التقليدية من الفهارس. وقد عرفت المنظمة الدولية للمقاييس والمعايير تعنى مارك "القالب المصمم لأغراض تبادل المعلومات " Format for Information Exchange (INEX). ويحمل المعيار رقم ISO 2709.

ظهر مارك في أبريل ١٩٦٦ وبدأت إجراءات برمجته وقد استدعى ذلك تحليل النظام وتصميم البرنامج وقد تم ربط هذا المشروع في وقتها بنوعية

واحدة من الوثائق وهي (الكتب) وبدأت عملية توزيع تسجيلات مارك في سبتمبر ١٩٦٦م، وتم إرسال أول شريط ممغنط يحتوي على تسجيلات مارك في أكتوبر ١٩٦٦م، تلاه بعد ذلك بأسبوعين فقط نوفمبر١٩٦٦م بدء الخدمات البريدية لتوزيع هذه التسجيلات[١].

وقد أدت الدراسات المتلاحقة إلى تحسين فاعليته وتسميته (مارك ٢) حيث يشمل وصفاً ببليوجرافياً لكافة أشكال مصادر المعلومات، وبعد إدخال التحسينات عليه أجمع مستخدموه على أنه أفضل نظام موجود يمكن استخدامه للوصف الببليوجرافي لكافة أوعية المعلومات من حيث الدقة والشمول وسهولة الاستخدام في المعالجة الفنية واسترجاع المعلومات، مما نتج عنه استخدامه في كثير من المكتبات البريطانية، وتقوم مكتبة الكونجرس منذ ذلك الوقت بإنتاج الأشرطة الممغنطة بموجب هذا النظام التي توزعها على المكتبات المشاركة أسبوعيًا.

تم التنسيق بين مكتبة الكونجرس و المكتبة الوطنية الكندية في عام ١٩٩٨م؛ لتوحيد صيغ مارك الأمريكي مع صيغة مارك الكندي، للوفاء بالاحتياجات الجديدة التي ظهرت في مجال الفهرسة. واتفقت المكتبتان على إصدار صيغة جديدة من نظام مارك تسمى مارك ٢١ لتناسب احتياجات القرن الحادي والعشرين. ويعتمد مارك ٢١ على المواصفة الأمريكية Z39.2 الخاصة بتبادل البيانات الببليوجرافية، وأيضاً يعتمد على المعيار الدولي الصادر عن الأيزو ISO 2709 والخاص بتبادل البيانات.

---

(١) علي بن شويش الشويش. مارك MARC والبيانات الخلفية (الميتاداتا) Metadata: علاقة ندية أم تكاملية؟، تاريخ الاستشهاد ٢٠/١١/٢٠٠٧، مسار
http://www.yamsyaf.com/informatics1.htm

لقد حقق هذا المعيار فائدة كبيرة للمكتبات ومرافق المعلومات، حيث مكنها من التحول إلى الشكل الآلي بما يحقق المرونة التي تسعى إليها نظم استرجاع المعلومات. إلى جانب هذا عمل معيار مارك على مشاطرة المصادر والتعاون الدولي بين المكتبات. ويمكن تلخيص أهمية معيار مارك في الملامح التالية:

١- يحقق معيار مارك إمكانات التعاون وتوفير المصادر وتبادلها عن بعد بين المكتبات المختلفة على اختلاف المسافات الجغرافية واختلاف أنواعها.

٢- تستطيع المكتبات المختلفة الانضمام إلى مشروعات الفهارس الموحدة، واستخدام قواعد البيانات على الخط المباشر.

٣- يمكن من خلال استخدام معيار مارك التحكم في إتاحة الفهارس الآلية المباشرة من خلال إدخال المعلومات نفسها مرة واحدة فقط.

٤- يوفر هذا الشكل إمكانات استخدام البيانات ذاتها في أكثر من نظام آلي على اختلاف مدد الاستخدام.

٥- يتصف النظام بدقة أكبر في إمكانات البحث، حيث يمكن البحث بأي حقل أو حقل فرعي.

٦- يلائم معيار مارك فهرسة جميع أوعية المعلومات التقليدية وغير التقليدية.

٧- يتيح إمكانية استخدام كافة الهجائيات الخاصة بكل اللغات الحية.

٨- يتوافق مع قواعد الفهرسة الأنجلوأمريكية وهي أفضل وأشهر قواعد على المستوى العالمي.

يعتمد معيار مارك في تكوين التسجيلات الببليوجرافية على مجموعة من المكونات تتدرج من الأكبر إلى الأصغر، ويمكن تمثيلها في الجدول التالي:

- التسجيلة الببليوجرافية :

وهي وحدة بناء الفهرس في معيار مارك وتتكون التسجيلة الواحدة لمصدر المعلومات من أنواع مختلفة من مجموعات الحقول؛ تختص المجموعة **الأولى** : الفاتح leader وتختص بكيفية إدخال التسجيلة. والمجموعة **الثانية** : الدليل directory ويتعلق بالمعلومات المدخلة ومحارف الحقول tages وأطوالها. أما المجموعة **الثالثة** : فهي الحقول المتغيرة variable fields التي تستقبل بيانات الوصف الببليوجرافي لمصادر المعلومات.

- الحقول:

يختص كل حقل ومن تسجيلات الفهرس في معيار مارك ببيان أو مجموعة من البيانات الببليوجرافية التي توضع في شكل حقول فرعية. ويبدأ كل حقل من الحقول بتاج الحقل ثم المؤشرات الدالة على طبيعة وبناء الحقل ثم علامات الحقول الفرعية وتسجيل البيانات الببليوجرافية بها.

- الحقول الفرعية:

هي البيانات المختلفة التي توجد وتسجل داخل حقل ببليوجرافي واحد مثل بيانات المسؤولية المختلفة في حقل العنوان وبيان المسؤولية، أو بيانات النشر في حقل النشر. ويأخذ كل حقل فرعي رمزًا أو تاجًا فرعيًا خاصًا به مثل **$ a**.

- التمثيلات:

وهي مجموعة الحروف الهجائية أو الأرقام الداخلة في حقول البيانات.

- المؤشرات :

وهي رقم أو رقمان توضع في بداية الحقل لتحديد خصائص الحقل.

وقد عمل هذا المعيار على جمع وتنظيم مختلف حقول الوصف التي تستخدم لوصف مصادر المعلومات، وذلك باستخدام الأرقام المتسلسلة المرتبة منطقيًا محارف أو تيجان MARC Tages لحقول الوصف، ثم التعبير عن بيانات كل حقل باستخدام مجموعة من الحقول الفرعية تعتمد على استخدام الحروف الهجائية. ومن ثم فإن الشكل العام لحقول الوصف الببليوجرافي لمعيار مارك يكون كالتالي:

- التاج ٠١٠ يدل على رقم الضبط الخاص بمكتبة الكونجرس LCCN)).
- التاج ٠٢٠ يدل على الترقيم الدولي للكتب ISBN)).
- التاج ١٠٠ يدل على اسم شخص كمدخل رئيسي مؤلف (author)
- التاج ٢٤٥ يدل على عنوان العمل (والذي يشتمل على العنوان والبيانات الأخرى للعنوان وبيان المسئولية) title .
- التاج ٢٥٠ يدل على بيانات الطبعة edition.
- التاج٢٦٠ يدل على بيانات النشر publication information .
- التاج ٣٠٠يدل على بيانات الوصف المادي physical description .
- التاج ٤٤٠ يدل على عنوان السلسلة series statement/added entry.
- التاج ٥٢٠ يدل على التبصرة الخاصة بالتعليق أو التلخيص annotation note.

- التاج ٦٥٠ يدل على رأس الموضوع subject heading topical.
- التاج ٧٠٠ يدل على اسم شخص كمدخل إضافي (مؤلف مشارك، محرر، موضح).
- التاج ٧١٠ يدل على المداخل الإضافية للأسماء المشاركة في المسئولية.
- التاج ٨٠٠ يدل على المداخل الإضافية للأسماء الشخصية والبيانات المتعلقة بها كتاريخ الميلاد والوفاة.
- التاج ٩٠٠ يدل على البيانات الإضافية التي ترغب المكتبة بإضافتها وتتعلق بملامح محلية الاستخدام.

٣- معيار: EAD

يعبر معيار EAD عن ترميز الوصف الأرشيفي EAD - (Encoded Archival Description)، ويعتبر هذا المعيار أحد محددات الوثائق في لغة الترميز المعيارية الممتدة (Extensible Markup Language (XML. ويدعم هذا المعيار إمكانية وصف الوثائق الرقمية ومعالجتها واسترجاعها في بيئة الويب. وقد خرج هذا المعيار بدعم من جمعية الأرشيفيين الأمريكية the Society of American Archivists. ويمتاز هذا المعيار بقدرات استرجاعية عالية تأتي من توفير عناصر للمعالجة تغطي مختلف أجزاء الوثيقة، فضلاً عن إمكانية البناء والعرض الهرمي للمعلومات الرقمية.

٤- معايير مصادر المعلومات المصورة :

تشكل مصادر المعلومات المصورة على الويب حجمًا هائلاً من الملفات الرقمية التي تحملها الويب مخزنة على خادمات الويب. ولذلك فقد خرجت

بعض المعايير التي تحدد عناصر الوصف والمعالجة للمعلومات الرقمية المصورة. ومنها ما يلي[١]:

أ- معيار DIG35 Specifications :

وضع هذا المعيار هيئة International Imaging Industry Association، وهي هيئة تجمع بين مجموعة من الشركات العاملة في مجال إنتاج الصور الرقمية مثل Kodak ، Agfa ، Canon... وغيرها. ويحتوي المعيار على مجموعة من الحقول الموحدة المستخدمة لوصف الصور الرقمية ويرمز له بـ I3A لتمثيل اسم الجهة القائمة على بنائه. وكان أول إصدار لهذا المعيار في الثامن عشر من يونيو عام ٢٠٠١ وذلك بإخراج الإصدارة ١,٠، ثم يلي ذلك خروج الإصدارة ١,١ والتي تعمل بالتوافق مع ترميز لغة الترميز المعيارية القابلة للتوسع XML Tags؛ وذلك لما لهذه اللغة من قدرة على توفير البيئة الملائمة لعمل حقول الوصف الميتاداتا وصلاحيتها لاستدعاء الصور الرقمية وعرضها بشكل فردي. ولقد حرصت هيئة الإصدار على بناء هذا المعيار بالشكل الذي يمكن معه مواجهة صعوبات التعامل مع ملفات الصور الرقمية وتنظيمها واسترجاعها، وهي الصعوبات التي أخذت في التزايد مع انطلاق عمل محركات بحث الصور في تنظيم وبحث الصور. كما عملت هيئة إصدار هذا المعيار على صيانته في إصداراته المختلفة ليتوافق مع المعايير الأخرى مثل ISO's JPEG-2000 and MPEG-7، بالإضافة إلى متابعة التحديث في لغة XML ولغات التركيز المعيارية الأخرى[٢].

---

(١) سيد ربيع سيد. المصدر السابق.

(٢) DIG35: Metadata Standard for Digital Images. XMLCoverPages , 2002 , cited at 21/6/2004 , cited at http://xml.coverpages.org/dig35.html.

ب- معيار VRA3 :

يتناول هذا المعيار the Visual Resource Association Core categories v3 (VRA 3)) ويبلغ عدد عناصره ١٧ عنصرًا ترتكز على أوعية المعلومات المصورة غير أنها ليست في نفس تعقيد معيار مارك Machine Readable Cataloguing MARC)). وتتعدى عناصر وصف هذا المعيار أوعية الصور الثابتة الرقمية إلى مختلف المعلومات المرئية الأخرى مثل فنون الرسم على الجدران وفنون الأشكال الهندسية إلى الصور المطبوعة[١].

ج- معيار Metadata for Long term Preservation

وضع هذا المعيار The Networked European Deposit (NEDLIB) Library وعرف باسم Metadata for Long term Preservation، ويتعلق بالعناصر والحقول الجوهرية المستخدمة في وصف المصادر الإلكترونية وخاصة الصور التي يتم حفظها لأمد طويل وجاء المعيار مركزًا على ثمانية عناصر رئيسة وثمانية وثلاثين عنصرًا فرعيًا.

١ / ٤ / ٢ / ٢ التنظيم وتحليل المحتوى :

لقد عملت الحاسبات الآلية وتطورها الدائم على توحيد مفاهيم مختلفة لمجموعة مصطلحات في مصطلح واحد، إلى ظهور مجموعة جديدة من مصطلحات نظم استرجاع المعلومات ارتبطت بالبيئة الرقمية. ويعكس واقع مصطلحات التحليل الموضوعي وتحليل المحتوى والتكشيف والمعالجة الفنية ما

---

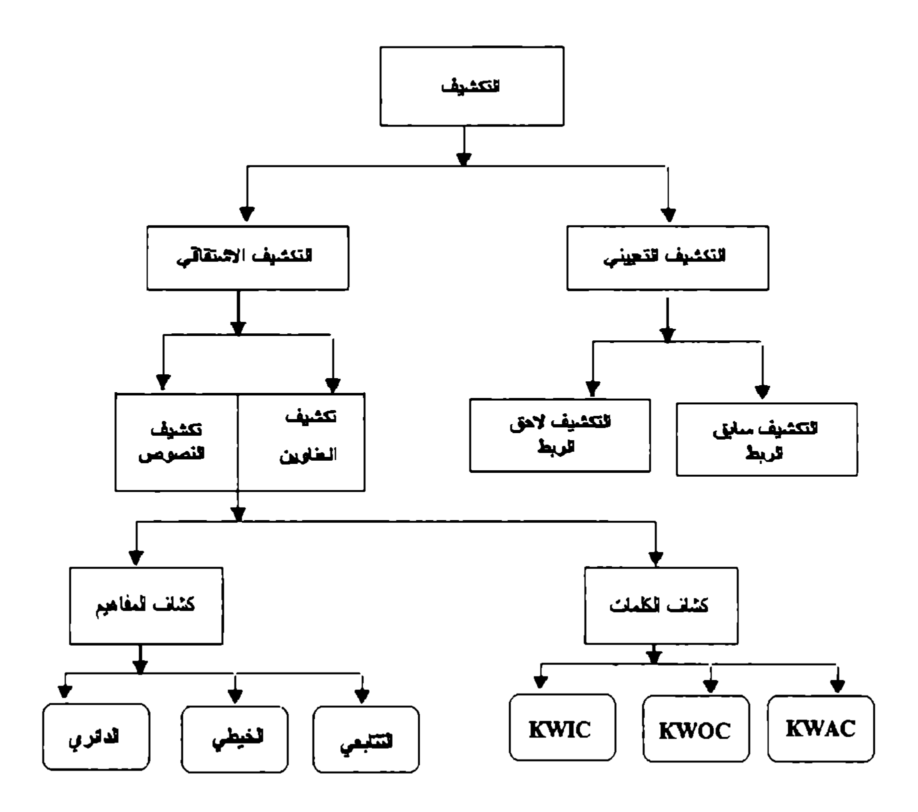
(١) Metadata: Image Description, Indexing and Cataloguing. tasi , 2004 , cited at 21/6/2004 , cited at http://www.tasi.ac.uk/advice/delivering/meta.html.

نعنيه بتداخل وامتزاج المفاهيم معًا؛ فلقد أضحى مصطلح تحليل المحتوى ينسحب على التحليل المعلوماتي لكلمات النصوص إلى جانب التحليل التقني الفني لأشكال مصادر المعلومات المصورة من حيث اللون والشكل والبنية. مثال آخر فلقد أصبح مصطلح التنظيم ينسحب على الوصف الفني والتحليل الموضوعي معًا، خلاف ما كان عليه من تحديد لمصطلح التحليل الموضوعي، حيث باتت إستراتيجية البحث تتكون من مختلف كلمات وحقول وصف التسجيلة البيليوجرافية سواء التي تصف الخصائص المادية والشكلية أم تلك التي تصف المحتوى والموضوع الداخلي لمصدر المعلومات. غير أننا في هذا الجزء سوف نركز الحديث عن عمليات تحليل المحتوى بالطرق الآلية اعتمادًا على عمليتي التكشيف والاستخلاص.

١ / ٤ / ٢ / ٢ / ١ التكشيف الآلي للمعلومات :

تتكامل عملية التكشيف مع مجموعة أخرى من عمليات التحليل الموضوعي؛ لتحليل وكشف الكلمات الدالة على موضوع أو محتوى الوثائق المختزنة في نظم الاسترجاع. ومن بين عمليات التحليل الموضوعي الأخرى التصنيف ورؤوس الموضوعات والاستخلاص، غير أنها تختلف معًا في طرائق العمل والنتيجة النهائية وشكل التعبير عن المحتوى سواء أكان رموزًا أم ألفاظًا أم فقرات. وتعتمد عملية التكشيف منذ بدايتها على استخدام أدوات المكانز في تحديد المصطلحات المقننة الملائمة للتعبير عن المحتوى ووثائق النظام. ومن ثم كان يرتبط التكشيف باستخدام اللغات المقيدة أو المضبوطة المعتمدة على تعيين مجموعة من المصطلحات الدالة على الموضوع. وقد تأثرت عملية التكشيف بإدخال وتطور الحاسب الآلي في المكتبات والمعلومات

كباقي العمليات الفنية الأخرى ؛ حيث أوجد الحاسب الآلي نوعًا جديدًا من التكشيف سمي بالتكشيف المبني على الاشتقاق يعتمد على استخدام الآليات الإحصائية أو آليات أخرى لتحديد أهم الكلمات المعبرة عن موضوع الوثيقة. ويوضح الجزء التالي أنواع التكشيف المستخدمة من قبل نظم الاسترجاع. ويوضح الشكل رقم (١ – ٦) أنواع الكشافات وعلاقاتها بلغات التكشيف الطبيعية والمقيدة[١].



الشكل رقم (١ – ٦) أنواع الكشافات

(١) تعريف التكشيف ونظمه تاريخ الاستشهاد ٢٠/١١/٢٠٠٧، متاح على:
www.jeeran.com/archive/2007/3/168846.html

❑ أنواع التكشيف الآلي:

ارتبط التعبير عن الموضوع دائمًا باستخدام الكلمات الدالة فقط دون الكلمات غير الدالة أو الأحرف. ومن ثم كانت الدلالة الموضوعية هي المعيار الأساس لاختيار الكلمات أو الواصفات الموضوعية، غير أن هذه الدلالة قد تكون الروح التي تسكن جسد الكلمة الواحدة أو جسد الكلمات المكونة للمفهوم. ومن هذا المنطلق كان نوعا إجراءات التكشيف متمثلين في التكشيف المبني على الكلمات والتكشيف المبني على المفهوم.

١- كشاف الكلمات:

يعمل هذا الكشاف لإخراج الكلمات الدالة على محتوى الوثيقة، وتستخدم الكشافات الآلية أساليب مختلفة في مثل الأساليب الإحصائية والاعتماد على موقع الكلمات واستخدام عناصر محددة من الميتاداتا لإخراج الكلمات الدالة مثل Meta key word. ومن الممكن أن يقتصر التكشيف على العنوان أو أن يمتد إلى كامل المتن الوارد في الوثيقة. ومن أمثلة هذه الكشافات ؛ كشاف الكلمات الدالة داخل وخارج السياق.

٢- كشاف المفاهيم :

إن هذا النوع من الكشافات يعتمد على التقاط المفاهيم وليس الكلمات، فالأصل هنا هو الحصول على المفهوم سواء كان ذلك متمثلا في كلمة واحدة أو مجموعة من الكلمات. وعلى ذلك تقوم عملية تكشيف المفاهيم الآلية على استبعاد مختلف الحروف والكلمات غير الدالة من العنوان أو النص المكشف والإبقاء على المفاهيم الرئيسية في النص. وكما وضح الشكل السابق، فإن هناك ثلاثة أنواع من كشافات المفاهيم هي كالتالي:

- **كشاف المفاهيم الخيطي** : ويعتمد على التبديل بين المفاهيم وبعضها البعض في الترتيب الهجائي للمداخل، بحيث يوضع المدخل الواحد على سطرين يكون المفهوم الداخل في الترتيب على السطر الأعلى وباقي المفاهيم الأخرى المكملة للعنوان على السطر الأسفل، ومن ثم تتبادل المفاهيم فيما بينها هذه الأماكن وفقًا لدور كل مفهوم على حدة.

- **كشاف المفاهيم التتابعي** : توجد المفاهيم المشتقة من عنوان واحد على سطر واحد وتبدأ في السير في اتجاه واحد بحيث يأخذ كل مفهوم دوره في الترتيب كمدخل للعنوان المكشف.

- **كشاف المفاهيم الدائري** : وفيه ترتب المفاهيم معًا ثم يأخذ كل مفهوم دوره كمدخل للعنوان المكشف وإذا ما حل موضعه مفهوم آخر في الترتيب، فإنه يتأخر إلى نهاية ترتيب المفاهيم معًا.

❑ طرق التكشيف في نظم المعلومات :

تنتهج عملية التكشيف منهجين مختلفين في الوصول إلى الكلمات أو المصطلحات الدالة المعبرة عن محتوى وموضوع وثائق نظام المعلومات. فإما أن تتم عملية التكشيف بتعيين المصطلحات أو الواصفات من خلال المكشف الإنسان، أو أن تتم من خلال استخدام الحاسبات الآلية وعندئذ يسمى بالتكشيف الاشتقاقي. ففي التكشيف التعييني يعتمد المكشف على تحديد موضوع ومفهوم الوثيقة ثم اللجوء إلى أدوات اللغة المقيدة من قوائم رؤوس موضوعات ومكانز لاختيار الملائم من الواصفات والمصطلحات الدالة على هذا المحتوى وتكون هي الكلمات الكشفية النهائية. أما في حالة استخدام

الحاسب الآلي، ولأن الحاسبات الآلية لا تملك الفهم والانتقاء كما في حالة المكشف الإنسان، فإن الحاسب الآلي يعمد إلى استخدام أساليب أغلبها إحصائية للوصول إلى أي من كلمات العنوان أو النص تصلح لتكون الكلمات الكشفية عن هذه الوثيقة. وفيما يلي تناول لأسلوبي التكشيف[1]:

أولاً : التكشيف بالتعيين:

يدل مصطلح التعيين assigned indexing هنا على إدخال كلمات أو واصفات قد لا تكون في سياق النص أو العنوان المكشف، غير أنها تعبر بدقة عن محتوى وموضوع هذه الوثيقة، مصدر هذه الكلمات أو الواصفات يأتي من أدوات العمل المستخدمة كقائمة رؤوس الموضوعات أو المكانز. وقد قصر استخدام الحاسب الآلي في هذا النوع واعتمد كلية على العنصر البشري، غير أن الباحث يرى أنه الآن مع تطور إمكانات الحاسبات الآلية وإمكانية معالجة المعاجم الآلية فإن الحاسب الآلي باستطاعته إجراء عمليات التكشيف بالتعيين وذلك على مرحلتين ؛ **الأولى** : القيام باشتقاق الكلمات الدالة من النص أو العنوان، **الثانية** : استبدال الكلمات المشتقة بمصطلحات مقابلة بحسب المعجم الآلي الموجود في قاعدة المعرفة لنظام استرجاع المعلومات. وقد تطور التكشيف بالتعيين على مرحلتين هما:

١ - الربط المسبق:

تتم عملية الربط بين المصطلحات في مرحلة التكشيف، وهو الأسلوب الملائم للاتباع في حالة الكشافات التقليدية المطبوعة، حيث إنه في حالة اتباع

---

(١) أحمد بدر و محمد فتحي عبد الهادي و ناريمان إسماعيل متولي . **التكشيف والاستخلاص** : دراسات في التحليل الموضوعي . - القاهرة: دار قباء، ٢٠٠١ .- ص٤٧ -٤٩.

الربط اللاحق مع هذا الشكل من أشكال الكشافات، فإنه يكون على المستفيد أن يجمع كماً كبيراً من التسجيلات الواردة تحت كل مصطلح من المصطلحات، ثم يقوم بفرزها بنفسه للتعرف إلى التسجيلات المتعلقة بالمصطلحات التي يبحث عنها مجتمعة، وهو أمر لاشك صعب ويحتاج إلى جهد ووقت كبيرين.

ويعد استخدام قوائم رؤوس الموضوعات ملائماً لهذا النوع من التكشيف حيث ترد المصطلحات في القائمة بشكل يمكن به أن تستقيم تلك المصطلحات بمفردها في القائمة، وأن تعبر عن المفاهيم المختلفة دون الحاجة إلى ربطها بمصطلحات أخرى كما هو الحال بالنسبة للمكانز.

٢ - الربط اللاحق:

يتم في هذا النوع من التكشيف الربط بين المصطلحات عند استرجاع الوثائق، أي في مرحلة البحث، وتستخدم عوامل المنطق البولياني للربط بين المصطلحات عند صياغة إستراتيجية البحث. وقد أصبحت معظم نظم الاسترجاع على الخط المباشر تعتمد على الربط اللاحق للمصطلحات.

ويقوم المكشف في مرحلة التكشيف بإعداد تسجيلة لكل مصطلح من المصطلحات الموضوعية، وتحديد أرقام الوثائق المرتبطة بالمصطلح في التسجيلة نفسها، وذلك بعد أن يتم ترميز الوثائق التي يضمها النظام بأرقام مسلسلة، وعند استرجاع الوثائق، فإن النظام يقوم بالبحث في تسجيلات المصطلحات.

ثانيًا : التكشيف الاشتقاقي:

يعتمد هذا النوع كلية على استخدام الحاسب الآلي في عمليات التكشيف، ومن ثم يقوم الحاسب باشتقاق مصطلحات التكشيف سواء من

نص الوثيقة، أم من المستخلص أو من العنوان، ويندرج تحت هذا النوع كل من تكشيف النصوص وتكشيف العناوين.

أ- تكشيف النصوص:

يتم وفقاً لهذا النوع من التكشيف الاعتماد على النص الكامل للوثيقة، ويحتاج المستفيدون في بعض الأحيان إلى هذا النوع من التكشيف، ومن ذلك على سبيل المثال في حالة تكشيف القوانين والدساتير.

ب- تكشيف العناوين:

يتم تحديد الكلمات المفتاحية من عنوان الوثيقة، فكل كلمة في العنوان هي بمثابة مصطلح تكشيف باستثناء الكلمات غير الدالة، وتعتمد فكرة هذا النوع من الكشافات على أن المؤلف يختار لوثيقته دائماً عنواناً معبراً عن المفاهيم التي تنطوي عليها تلك الوثيقة، ولعل مما يعاب على هذا التكشيف أن بعض المؤلفين يختارون في بعض الأحيان عناوين مضللة لا توحي بمضمون الوثيقة، ومن بين كشافات العناوين نذكر الآتي:[١]

١- كشاف الكلمات المفتاحية الدالة في السياق Keyword In Context (KWIC): ترتب الكلمات الدالة هجائياً ويأتي بعدها باقي كلمات العنوان تبعاً لعدد الكلمات الواردة فيه من الكلمات غير الدالة.

٢- كشاف الكلمات المفتاحية الدالة خارج السياق Keyword out of Context (KWOC) : تؤخذ الكلمات من مكانها في العنوان وتوضع في الهامش الأيسر عند تكشيف العناوين الإنجليزية، والهامش الأيمن عند تكشيف العناوين العربية.

---

(١) أحمد بدر ومحمد فتحي عبد الهادي وناريمان إسماعيل متولي. المصدر السابق نفسه.

٣- كشاف الكلمات المفتاحية الدالة بمحاذاة السياق Keyword Augmented Context (KWAC) يتم إضافة مصطلحات من خارج العنوان لإيضاح المفهوم ودعم الجانب الدلالي للعنوان.

ثالثًا : إجراءات وخطوات التكشيف الآلي :

تقوم نظم استرجاع المعلومات على الويب بإجراء عمليات التكشيف على الصفحات التي يضيفها برنامج الزاحف، وذلك لاستخراج الكلمات المفتاحية الدالة على محتوى هذه الصفحات والتعبير عن المحتوى الداخلي ثم تسجيلها في ملف الكشاف. وهذه الإجراءات تجرى على النحو التالي:

١- إعداد قائمة الاستبعاد للكلمات كثيرة التواجد غير الدالة وقد عرفها لون على أنها :

- هي مفردات لا تحمل مضمون أو معاني دلالية مستقلة.
- وقد تكون حروفًا أو أدوات لازمة لتركيب الكلام والربط بين أجزائه.
- صفات أو نعوت لمفاهيم أخرى.
- كلمات عامة وغير مخصصة لموضوع من الموضوعات.
- قد تكون مفردات لا يتوقع السؤال عنها من جانب المستفيد.

٢- يتم تجريد الكلمات من السوابق واللواحق الواردة معها.

٣- حساب عدد مرات التكرار لكل الكلمات وترتيبها تربيًا تنازليًا بحسب عدد مرات التكرار، ثم اختيار الكلمات التي تتصدر الترتيب.

٤- يتم اعتبار الكلمات التي تجاوز عدد مرات تكرارها عدد (س) من المرات، كلمات دالة أو مفتاحية لمحتوى الوثيقة داخل نظام الاسترجاع.

١ / ٤ / ٢ / ٢ / ٢ الاستخلاص الآلي للمعلومات:

تعمل المستخلصات على صياغة الملامح الموضوعية للوثيقة في شكل فقرات تتيح للمستفيد التعرف إلى المحتوى. وهي أحد أشكال التحليل الموضوعي مع التصنيف والتكشيف ورؤوس الموضوعات، ويمكن القول إن المستخلصات قد خرجت لأداء وظائف محددة ذكرها أحمد بدر وهي :

١- تيسير وتسهيل عمليات الاختيار، أي أنها تحدد للمستفيد طبيعة الموضوع أو المحتوى للوثيقة.

٢- تتيح للقارئ التعرف إلى محتوى الوثائق مع اختصار الوقت والجهد في الوصول إلى الوثائق ذاتها وقراءتها.

٣- التعرف إلى محتويات الوثائق في الموضوع التي قد كتبت بلغات غير لغة المستفيد.

٤- إمكانية التعرف إلى الإنتاج الفكري المنشور حديثًا دون الانتظار لوصول الوثائق ذاتها، بما يضمن التحديث الدائم للمعرفة في المجال.

٥- تقدم المستخلصات الدعم لبعض خدمات المعلومات مثل الإحاطة الجارية والبث الانتقائي للمعلومات.

تعلو عملية الاستخلاص عملية التكشيف في التعبير والدلالة على المحتوى الموضوعي للوثائق ؛ حيث تملك المستخلصات قدرات أكبر في التعريف

بالمحتوى من خلال عرضه في شكل فقرة أو فقرات متعددة على اختلاف المستخلصات المختصرة أو الموسعة. وتعتمد المستخلصات الآلية اعتمادًا كليًا على استخدام الحاسب الآلي في التعرف إلى الكلمات الدالة التي تحدد بدورها أكثر الجمل في النص دلالة وتعبيرًا عن محتوى الوثيقة. وليست الطريقة الإحصائية هي وحدها التي يعتمد عليها الحاسب الآلي في إجراء عملية الاستخلاص الآلي لوثائق نظام استرجاع المعلومات، وإنما هناك طرائق أخرى يمكن التعريف بهام فيما يلي[1]:

الطريقة الأولى : بيتر لون (إحصائية):

يعتمد بيتر لون في مذهبه للاستخلاص الآلي على الإحصاء والتكرار ؛ حيث يرى أن الكلمات الأكثر أهمية في التعبير عن الموضوع هي الأكثر تكرارا في سياق النص. ويقوم الاستخلاص الآلي في هذه الطريقة تبعًا للخطوات التالية:

١- استبعاد جميع الكلمات غير الدالة من النص، باستخدام قوائم التوقف والمرور stop / go list.

٢- إحصاء تكرار ورود الكلمات التي لم تشملها قائمة التوقف، ثم ترتيب هذه الكلمات ترتيبًا تنازليًا وفقًا لتكرارها داخل نص الوثيقة.

٣- تحديد الكلمات التي يتكرر ورودها أكثر من عدد (س) من المرات، واعتمادها كلمات دالة على محتوى الوثيقة.

---

(١) هاني محيى الدين عطية. **محاضرات غير منشورة**.

٤- تحديد التجمعات التي تشتمل على عدد من الكلمات عالية التكرار في الجملة الواحدة.

٥- حساب معامل أهمية كل جملة وفقًا لعدد الكلمات الدالة في كل تجمع للجملة ثم جمع أهمية التجمعات معًا للخروج بمعامل أهمية الجملة كاملة، ويتم ذلك بتحديد عدد كلمات الجملة (w) ثم حساب عدد الكلمات الدالة (w1) ويعبر عن ذلك بالمعادلة التالية:

معامل أهمية الجملة = عدد الكلمات الدالة(w1)\ عدد كلمات الجملة (w)

٦- ترتب الجمل التي تحتوي على الكلمات الدالة وفقًا لمعامل أهميتها ترتيبًا تنازليًا ثم التقاط الجمل الأكثر أهمية وطباعتها وفقًا لتسلسلها في الوثيقة.

الطريقة الثانية : طريقة بيس (أهمية موقع الكلمات):

يرى بيس أن موقع الكلمات الواردة في الوثيقة يدل دلالة كبيرة على أهمية أو عدم أهمية هذه الكلمات في الدلالة على المحتوى. ومعايير هذه الرؤية هي :

أ- إن الكلمات المفتاحية هي تلك التي ترد في الجملة باعتبارها كلمات مفتاحية أو الواردة في عنوان الوثيقة أو الواردة في عناوين الجداول والأشكال التوضيحية.

ب- إن الجمل التي ترد في مواقع مثل بداية الوثيقة أو نهايتها أو في الخاتمة أو بداية الفقرات هي الجمل التي لها دلالة أكبر في موضوع الوثيقة.

ج- تعطي طريقة عرض الكلمات والجمل دلالة على أهمية أو عدم أهمية هذه الجمل في التعبير عن الموضوع؛ حيث تكتب بعض الكلمات بخط بارز أو لون مخالف ما يحمل معه دلالة على أهمية هذه الجمل في السياق والمحتوى.

د- هناك بعض العبارات التي تشير إلى أهمية ما يرد بعدها من كلمات مثل عبارات : (لقد كان الهدف من هذا البحث.....)

الطريقة الثالثة : طريقة راش وزملائه (أهمية السياق) :

يعتمد راش في استخراج الجمل الدالة لصياغة المستخلص على أن السياق ذاته هو المحدد والدال على الجمل الأكثر أهمية. فدائما ما يلجأ المؤلف إلى عرض موضوع الوثيقة باستخدام نماذج وأشكال من العبارات المختلفة، نوضحها فيما يلي:

أ- عمل قائمتين للتحكم في المفردات، إحداهما للتوقف عن العبارات التي إذا وجدت فهي تشير إلى جمل يتم رفضها، والأخرى التي إذا وجدت فتشير إلى أهمية وقبول ما بعدها من جمل. وتشمل قائمة القبول كل ما يرد للتعبير عن محتوى أو موضوع الدراسة كنماذج (هذا البحث....، هذه الدراسة... ، البحث التالي....)

ب- مضاهاة نص الوثيقة مع قائمتي التوقف والقبول.

ج- حذف العبارات الاعتراضية غير المهمة.

د- قبول الجمل المتعددة المرتبطة التي تعبر عن أكثر من مفهوم إلى حد ثلاث جمل.

هـ- تعديل الجمل التي تم انتقاؤها ثم ترتيبها وطباعتها.

الطريقة الرابعة : استخدام النظم الخبيرة :

حاولت النظم الخبيرة الاعتماد على المناهج السابقة للوصول إلى أفضل طرائق الاستخلاص الآلي، وقد وضعت معايير لذلك تتمثل في: -

أ- أن يتعرف النظام إلى جميع النقاط المهمة من كلمات وجمل داخل النص باستخدام أي من الطرق السابقة.

ب- أن يتم الربط بين الجمل التي ترتبط بعلاقات موضوعية سواء التي وردت في المقدمة أم الخاتمة أم في أجزاء متفرقة من النص.

ج- يتم الوقوف على العبارات المتكررة في المستخلص المبدئي واستبعادها.

د- ترتيب الجمل الواردة في المستخلص المبدئي كي يستقيم المعنى.

هـ- أن يتيح النظام الخبير للعنصر البشري إجراء التعديلات والتغييرات اللازمة على المستخلص للوقوف على شكله النهائي.

تصل المستخلصات على اختلاف طرائق تكوينها إلى الشكل النهائي المكون من الإشارة الببليوجرافية الواصفة للوثيقة، بجانب الفقرة أو الفقرات المكونة للمستخلص الدالة على المحتوى. ومن ثم فإن المستخلصات تتركز في شكلين أساسيين هما المستخلصات الإعلامية والمستخلصات الشارحة. وقد أورد أحمد بدر نماذج أخرى متباينة من المستخلصات مثل:

- المستخلصات الشارحة .
- المستخلصات الإعلامية .

- المستخلصات الكشفية .
- المستخلصات النقدية .
- المستخلصات ذات الأغراض الخاصة .

١ / ٤ / ٣ المخرجات :

تمثل المخرجات في نظم استرجاع المعلومات الناتج النهائي لتشغيل وبناء النظام. وتأخذ المخرجات في نظم استرجاع المعلومات أشكالا مختلفة، لا تقف فقط عند المعلومات المسترجعة وإنما تتعدى ذلك إلى خدمات المعلومات المختلفة التي يؤديها النظام، إلى جانب ذلك فيمكن للنظام تقديم خدمات أخرى مثل التحديث والتطوير في العناصر المختلفة لبناء نظم الاسترجاع. وتكمن أهمية المخرجات هنا في كونها الشاهد الفعلي على نجاح أو فشل عمليات بحث واستدعاء المعلومات من النظام، فعوامل الاستدعاء والتحقيق لا تطبق في مرحلة الإضافة أو العمليات الفنية، وإنما يتم العمل بها في مرحلة استرجاع المعلومات. فيعكس معامل الاستدعاء مدى القدرة على استرجاع المعلومات المرتبطة بموضوع البحث، ويعكس معامل التحقيق مدى قدرة النظام على فهم موضوع البحث واسترجاع المعلومات الملائمة له دون عرض المعلومات التي تمثل شوشرة للمستفيد في أثناء الاسترجاع.

يستدعي نظام الاسترجاع مجموعة الوحدات أو التسجيلات ردًا على إستراتيجية البحث المصاغة من جانب المستفيد، ومن ثم فإنه على المستفيد تحري الدقة في اختيار كلمات البحث المعبرة عن حاجاته من المعلومات. ويفترض بهذه التسجيلات المسترجعة أن تكون موافقة لموضوع المستفيد، غير

أنه يمكن القول إن نظام استرجاع المعلومات قد يتحمل أخطاء صياغة مفردات البحث من جانب المستفيد ؛ حيث يملك المستفيد وضع الكلمات البحثية بنفسه كما يملك الحكم على المعلومات والنتائج المسترجعة بذاته. ومن ثم يكون معيارا الاستدعاء والتحقيق هما معاملي قياس كفاءة المستفيد وليس النظام. ولتوضيح ذلك فإن كلاً من المستفيد أو وسيط البحث يقومان باختيار الكلمات المفتاحية ومعاملات الربط بينها، ثم إدخالها إلى نظام الاسترجاع ومن ثم يستجيب النظام لترتيب ونوع الربط بين الكلمات المفتاحية جميعًا؛ متعرفًا عليها كاستراتيجية بحث تستدعي معلومات محددة يختارها من بين المعلومات الأخرى. ثم يكون تطلع المستفيد للنتائج المسترجعة ليرى مدى إيفاء المعلومات المسترجعة بتلبية حاجاته الموضوعية. إن ثمة أمرًا يجب الانتباه إليه في معايير الحكم على كفاءة المخرجات ألا وهو إن نظم استرجاع المعلومات الرقمية عامة والتي تستخدم اللغة الطبيعية خاصة كمحركات بحث الويب، تحتاج إلى معاملات أخرى ومنهج آخر للحكم على كفاءة النتائج المسترجعة، وهذا يرجع إلى الأسباب الآتية:

- تعمل نظم استرجاع المعلومات الآن في بيئة تعتمد أكثر ما تعتمد على صياغة الوثائق في المجال الموضوعي بألفاظ ومصطلحات مختلفة قد يستخدمها أو لا يستخدمها المستفيد عند استرجاع المعلومات.
- أن نظم استرجاع المعلومات على الويب تستخدم لغة التكشيف الطبيعية المعتمدة على التكشيف بالاشتقاق كأسلوب لتنظيم الوثائق، ومن ثم تصبح مختلف كلمات الوثيقة القريبة للموضوع والبعيدة عنه مداخل تستدعي هذه الوثيقة في حالة البحث بأي من هذه الكلمات.

- تتسم معاملات الاستدعاء والتحقيق بالكفاءة ودقة التطبيق فقط في حالة الاستخدام مع نظم الاسترجاع التي تستخدم اللغة المقيدة المحكومة بقوائم رؤوس الموضوعات والمكانز.
- إن بيئة الويب بيئة ديناميكية المحتوى وعندما يقدم المستفيد على بحث أي من نظم استرجاع المعلومات على الإنترنت فهو يتعامل مع صندوق أسود من المعلومات لا يستطيع تحديد ملامح الوثائق داخله.
- مدى وضوح الحاجة المعلوماتية في ذهن المستفيد، ومدى قدرته في الحكم على النتائج المسترجعة.

إن ثمة مستويات مختلفة لقياس المعلومات المسترجعة توجد في ذهن المستفيد عند تطلعه للنتائج المسترجعة. هذه المستويات تحتوي ضمنًا على استخدام معاملي الاستدعاء والتحقيق. وهي ثلاثة مستويات تتحدد في:

١- بعض وحدات النتائج المسترجعة يتطابق تمامًا وحاجة المستفيد من المعلومات. وهو ما يمثله التحقيق.

٢- بعض وحدات النتائج المسترجعة يقترب أو يبعد قليلا من حاجة المستفيد من المعلومات.

٣- بعض وحدات النتائج المسترجعة لا تتصل من قريب أو بعيد بحاجة المستفيد من المعلومات. وهو ما تمثله الشوشرة في نظم الاسترجاع.

لقد ذكر طلال الزهيري أن هناك مجموعة من المشكلات التي يتوقع أن يكون لها تأثير على كفاءة النتائج المسترجعة وهي ذات صلة مباشرة بدور المستفيد في اختيار وحدات المعلومات من المخرجات، مثل ما يأتي[١].

---

(١) **دور المستفيد في كفاءة استرجاع المعلومات**. طلال ناظم الزهيري، تاريخ الاستشهاد ٢٠٠٧/١١/١١، مسار المصدر http://azuhairi.jeeran.com/files/81727.doc

١- أحكام المستفيد على نتائج عملية الاسترجاع تخضع لدوافع شخصية غير معلنة.

٢- في قواعد البيانات الببليوجرافية التي تقدم مستخلصات البحوث فقط قد لا يتمكن المستفيد، خاصة قليل الخبرة من اكتشاف التسجيلات الملائمة لقلة المعلومات المتوافرة في المستخلص .

٣- عندما تختلف لغة قاعدة البيانات عن لغة المستفيد تصبح عائقاً أمامه للحكم بدقة على مدى صلاحية التسجيلات المسترجعة .

٤- ضيق وقت المستفيد في الغالب لا يمكنه من استطلاع التسجيلات المسترجعة كافة خاصة عندما يكون عددها كبيرًا إذ يعمل على الإسراع في عملية الاستطلاع أو تجاهل استطلاع عدد من التسجيلات.

٥- في بعض الأحيان يكون لدى المستفيد قرار مسبق بعدد التسجيلات التي يرغب في الحصول عليها.

٦- الكلفة المادية لعمليات البحث قد تحدد المستفيد أحيانًا بقبول عدد محدود من التسجيلات المسترجعة والاستغناء عن مجموعة أخرى حتى مع إقراره بملاءمتها.

١ / ٤ / ٤ الرقابة :

يحتاج نظام استرجاع المعلومات كغيره من النظم الأخرى إلى عنصر الرقابة على الأداء والتشغيل. وذلك للتعرف إلى جوانب القوة والضعف التي يتسم بها نظام استرجاع المعلومات، وعادة ما يقوم مديرو النظم بمتابعة تدفق العمليات داخل النظام والتعرف إلى الأخطاء التي تواجهها العناصر الفرعية المؤدية للمهام، ولا يقف الأمر عند ذلك، بل يتعدى إلى الحرص الدائم في

تطوير نظام استرجاع المعلومات سواء أكان من حيث المعلومات المختزنة وأشكالها أم من حيث تقديم استراتيجيات ولغات بحث تيسر للمستفيدين الوصول إلى المعلومات المحددة ورفع درجة التحقيق المطلوبة.

وقد لا تقتصر الرقابة على رأس نظام استرجاع المعلومات، وإنما تكون من خلال الوسيط الذي يحاكي الإجراءات المستمرة لعمل نظام استرجاع المعلومات. ولا تتم عملية الرقابة بهدف السيطرة على عمليات نظم استرجاع المعلومات فقط، وإنما يترتب على الرقابة عمليات التقويم والتقييم الدائمة المتتابعة للنظام.

١ / ٤ / ٥ التغذية المرتدة :

تمثل التغذية المرتدة أو التلقيم المرتد feedback لنظام الاسترجاع المرآة التي تنعكس عليها كفاءة وأداء النظام. فإن المستفيد عند تحديده لأكثر من نتيجة في صفحة النتائج، فهو بمثابة إعلام للنظام عن النتائج الصائبة وترك الخاطئة. كما أن إعادة عملية البحث أو استخدام مصطلحات أخرى هو إشارة للنظام عن عدم كفاية المعلومات المسترجعة من قاعدة البيانات.

ومن أهم الدراسات التي تناولت التغذية المرتدة كأحد مقاييس الكفاءة، دراسة Ryen W. White التي ركزت على سبل دعم التفاعل بين المستفيد والنظام من خلال دراسة بعض الأساليب الخاصة بالتلقيم المرتد أو التغذية المرتدة المرتبطة بالنتائج Relevance feedback (RF) technique. حيث طرح الباحث بعضًا من الواجهات التي ترتب نتائج البحث بأشكال مختلفة اعتمادا على آليات إحصائية وهرمية لترتيب النتائج المسترجعة، ويستطيع المستفيد استخدام هذه الآليات لطرح وجهات نظره في كفاءة نظام الاسترجاع في استدعاء الوثائق المرتبطة بموضوع البحث. ومن ثم تختلف هذه الأطروحة عن

دراستنا هذه في تناول نظم استرجاع المعلومات من جانب التقييم وقياس الكفاءة فقط، خلاف ما نهدف إليه من تناول نظم استرجاع المعلومات الرقمية مفهومًا وبنيةً وأداءً ثم تناول نظم استرجاع الويب من أكثر من جانب، بالإضافة إلى الحديث عن تقييم النظم وقياس الكفاءة من خلال معايير وعناصر تقييم شاملة(١).

وبعد الحديث السابق عن المكونات والعمليات الداخلية لنظام الاسترجاع، فإن ثمة موارد أخرى يجب أن يستند إليها نظام استرجاع المعلومات في القيام بمهامه وتطويرها في مراحل متعاقبة بما يحقق الكفاءة المرجوة من جانب المستفيدين. وهذه الموارد قد وضحها James A. فيما يلي(٢) :

١- الموارد البشرية : اختصاصيو التشغيل والمبرمجون ومصممو الصفحات والقائمون على تقييم النظام واختباره والوسطاء.

٢- موارد الأجهزة : أجهزة الحاسبات والماسحات الضوئية، وأجهزة الطابعات... وغيرها.

٣- موارد البرمجيات : برامج نظام التشغيل وبرامج نظام إدارة قواعد البيانات.

٤- موارد البيانات : المعلومات والتسجيلات وملفات المستخدم ومخزون قاعدة البيانات .

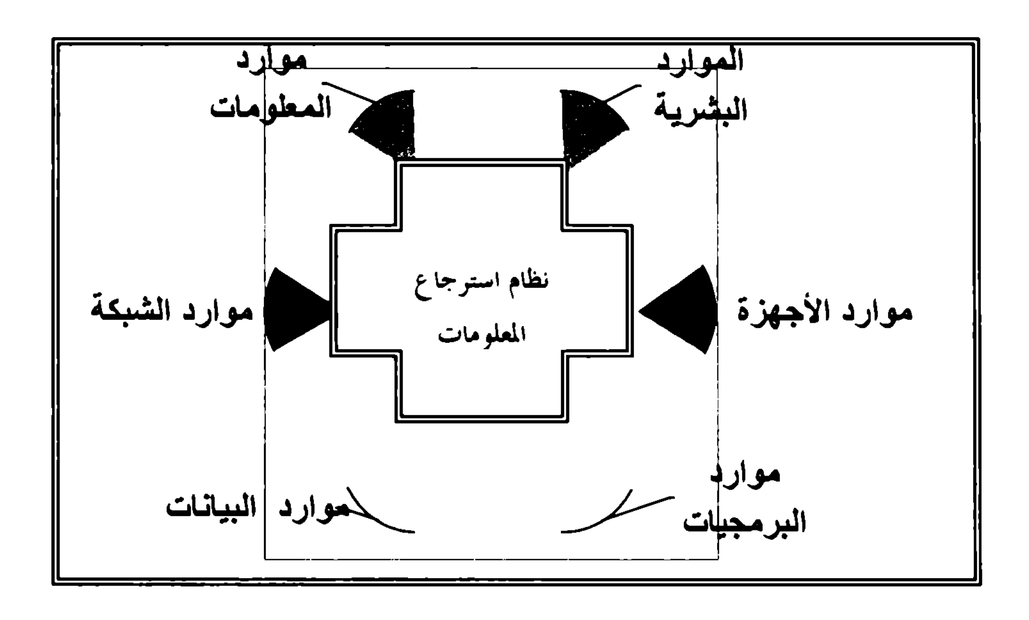
(١) White, Ryen W. Implicit Feedback for Interactive Information Retrieval, University of Glasgow, 2004, (doctoral thesis), cited at:
https://research.microsoft.com/~ryenw/thesis/thesis-files/RyenWhiteThesis.pdf

(٢) James A. O'Brien. Introduction to Information Systems: Essential for the Internet worked Enterprise. 9th.ed. NY.: McGraw-Hill, 2000, P.21.

٥- موارد الشبكة : أدوات الاتصال ومعالجات الاتصال وبرامج السيطرة ووسائل الوصول إلى الشبكة .

٦- موارد المعلومات : الكتب و الدوريات وغيرها من المصادر الورقية في النظم التقليدية.

وقد وضح طلال الزهيري هذه الموارد متمثلة في الشكل رقم (١ – ٧)[١]:



الشكل رقم (١ – ٧) الموارد التي يقوم عليها نظام استرجاع المعلومات

(١) طلال ناظم الزهيري. **العوامل المتصلة بنظام استرجاع المعلومات وتأثيرها على كفاءة الاسترجاع**، تاريخ الاستشهاد ٢٠٠٧/١١/١١، مسار المصدر
http://azuhairi.jeeran.com/files/81729.doc

١ / ٥ الخاتمة:

عملت الويب بآلياتها وتقنياتها على إصباغ نظم استرجاع المعلومات بالكثير من الملامح المستحدثة. وأهم هذه الملامح هي البعد عن دور البيانات الببليوجرافية كوسيط بين المستفيد والمعلومات النهائية، بل إن الأمر امتد إلى تحقيق المعالجة والتنظيم والبحث داخل وثائق النص الكامل. هذا إلى جانب اختلاف عمليات المعالجة والتنظيم للمعلومات الرقمية التي غلبت مبدأ التنظيم وتحليل المحتوى على عملية الوصف الفني والتحليل المادي. وقد وضح هذا الفصل أيضًا التنوع بين نظم استرجاع المعلومات وهو ما عكسه تقسيم نظم استرجاع المعلومات في قلب هذا الفصل. حيث يعكس ذلك التغيير الشكلي والموضوعي لنظم الاسترجاع والمعلومات المختزنة داخلها.

الفصل الثاني

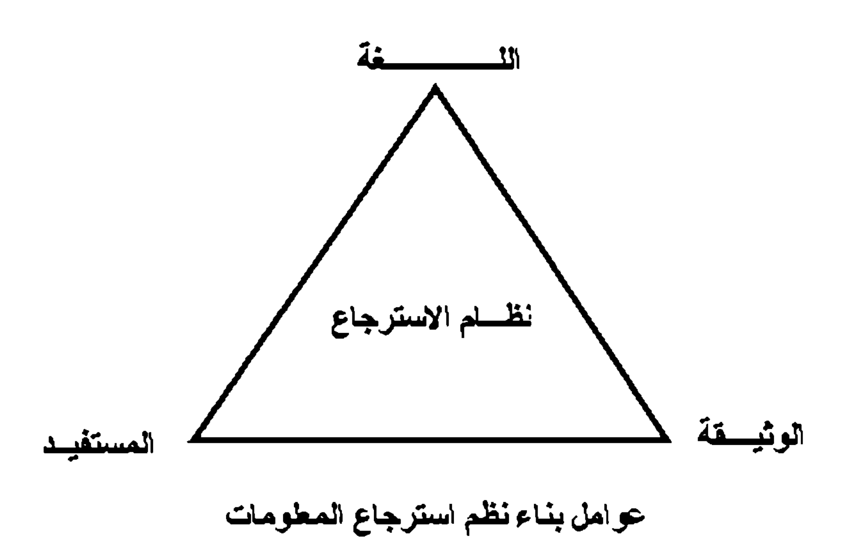
عوامل بناء نظم الاسترجاع

٢ /٠ التمهيد :

يتكون نظام استرجاع المعلومات من مجموعة من المكونات التي سبق الحديث عنها مثل المدخلات والمخرجات والعمليات الفنية، غير أن هذه المكونات لا تستطيع بمفردها بناء نظام استرجاع للمعلومات، ومن ثم فإن هناك عوامل أخرى تدعم بناء نظام الاسترجاع. تتمثل هذه العناصر في اللغة المتطلبة للحوار والتفاعل بين المستفيد والنظام، أيضًا المستفيد ذاته الذي قام وبني النظام من أجله لتلبية حاجاته من المعلومات، وأخيرًا الوثيقة التي بدونها لا يمكن للنظام أن يكون نظام استرجاع للمعلومات، وهذه الوثيقة لا تقتصر على مستوى أو شكل محدد من المعلومات وإنما تتسع لتشمل مختلف أنواع المعلومات التي يتم اختزانها في قواعد نظم استرجاع المعلومات. ويركز هذا الفصل على دراسة مفصلة لخصائص وملامح كل عنصر أو عامل من عوامل بناء نظام استرجاع المعلومات على حدة.

٢ / ١ عوامل بناء نظم الاسترجاع :

يتكون نظام استرجاع المعلومات التقليدي أو الرقمي من ثلاثة أضلاع هي: اللغة والمستفيد والوثيقة أو المعلومات التي يحويها النظام. ولا يمكن لنظام استرجاع المعلومات العمل في غياب أي من هذه الأضلاع، فنظام استرجاع للمعلومات بدون مستفيد مثلا ليس له قيمة، كما أن نظام استرجاع معلومات بدون وحدات أو تسجيلات في قواعد بيانات لن يقوم بمهمة الاسترجاع. ومن ثم كانت أهمية كل عنصر من العناصر الثلاثة المكونة لنظام الاسترجاع. يوضح الشكل رقم (٢-١) تكامل العناصر الثلاثة معًا.



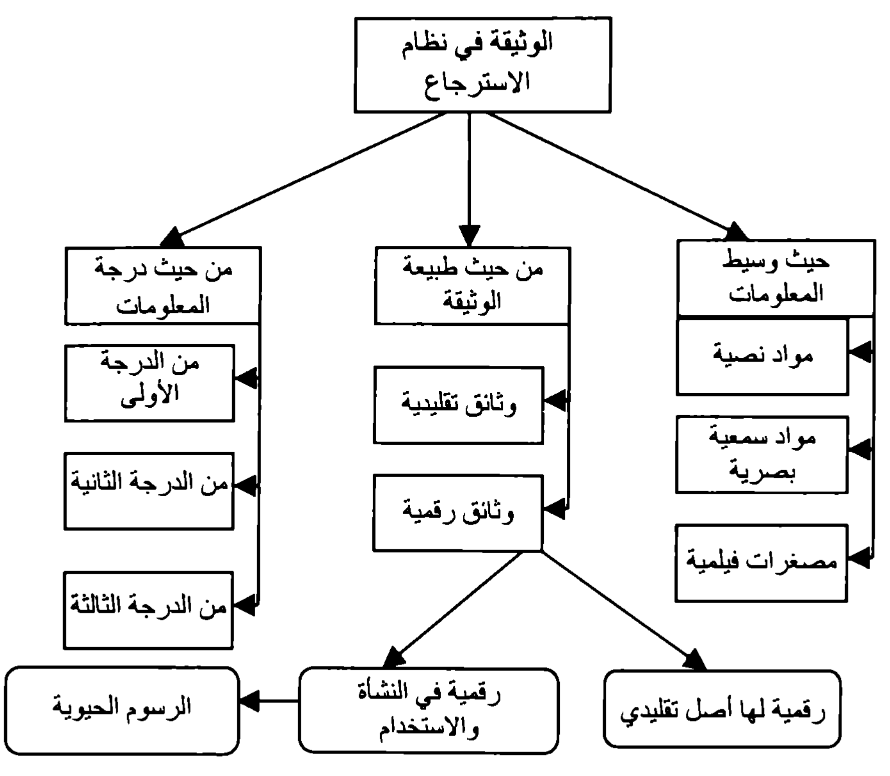
الشكل رقم (٢ -١) عناصر بناء نظم الاسترجاع

يوضح الشكل رقم (٢-١) العناصر التالية:

٢ / ١ / ١ الوثيقة:

يتسع مفهوم الوثيقة في نظم استرجاع المعلومات؛ ليشمل الكثير من أنواع وأشكال المعلومات المختزنة والمعالجة في نظم استرجاع المعلومات. فالكتاب في مرافق المكتبات التقليدية هو وثيقة داخل نظام الاسترجاع، والتسجيلة الببليوجرافية هي وثيقة في قاعدة بيانات مليزرة أو على الخط المباشر، وتلك التي يصاحبها مستخلص، والنصوص الكاملة داخل نظم استرجاع المصادر الأولية هي وثائق في نظام الاسترجاع، أيضًا كل من ملفات الصوت والصورة على اختلاف أنواعها هي وثائق داخل نظم استرجاع المعلومات. ويأخذنا الحديث عن أشكال المعلومات التي تقتنيها نظم استرجاع المعلومات إلى

محاولة التعرف إلى التقسيم العام للوثائق التي قد يحويها أي من نظم الاسترجاع في بيئتيه التقليدية والرقمية، وأنواع هذه الوثائق يوضحها الشكل رقم (٢-٢):



الشكل رقم (٢-٢) يوضح أنواع ومستويات الوثائق داخل نظم استرجاع المعلومات

يوضح الشكل رقم (٢-٢) أنواع الوثائق في نظم استرجاع المعلومات، والتي يبنى على إثرها واقع وعمل نظام استرجاع المعلومات:

٢ / ١ / ١ / ١ الوثائق من حيث وسيط المعلومات:

تحتوي نظم استرجاع المعلومات على وثائق ذات وسائط مختلفة بين المواد النصية والوسائط الأخرى.

أولاً : المواد النصية :

تعتمد المواد النصية على استخدام الكلمات والجمل لنقل المعلومات عبر الوثيقة وبثها عبر نظم استرجاع المعلومات. وقد ارتبط النص منذ البداية بالمصادر الورقية المطبوعة، غير أن النص قد انتقل إلى البيئة الرقمية دون اختلاف في شكل أو طبيعة التعامل مع الوثائق النصية. ويكمن وجه الاختلاف فقط بين المواد النصية في البيئة التقليدية والبيئة الرقمية في كون النص أيسر معالجة وأدق في الاسترجاع في البيئة الرقمية عنها في البيئة التقليدية.

ثانيًا : المواد السمعية البصرية :

تعتمد المواد السمعية البصرية على وسيطي الصوت والصورة سواء أكان ذلك في شكل منفصل في المواد الصوتية أو المواد المصورة أم ارتباطهما معًا في المواد السمعية البصرية. ويمكن القول إن البيئة الرقمية ونظم استرجاع الويب قد أوجدا حظًا أوفر للوسائط المتعددة على حمل المعلومات في عمليات المعالجة والتنظيم والتحليل الفني والاسترجاع مرة أخرى، وهو ما لم تجده متوفرًا في البيئة التقليدية. فاتجهت معظم محركات البحث على الويب إلى التخصص في التعامل مع ملفات الصوت والصورة أو ملفات الصور المتحركة.

ثالثًا : مواد المصغرات الفلمية :

قد لا نكون مبالغين إذا ما قلنا إن مواد المصغرات الفلمية قد انتهى الدور المهم الذي أدته باعتبارها وثائق لحمل المعلومات داخل نظم استرجاع المعلومات ؛ وهذا قد جاء نتيجة للتطورات الهائلة في إمكانات التصغير minimization

الذي هدفت له المصغرات الفلمية من توفير حيز اختزان مجموعات كبيرة من المعلومات. ولا يوجد للمصغرات الفلمية انعكاس في بيئة الويب الآن.

٢ / ١ / ١ / ٢ من حيث طبيعة الوثائق:

أولاً : الوثائق التقليدية :

يشمل مفهوم المواد التقليدية مختلف وثائق حمل المعلومات التي توجد في الشكل المطبوع، وهو ما بدأت به نظم استرجاع المعلومات. ولم تتطور الوثائق التقليدية ويتغير حالها مع نظم المعلومات سوى أنها قد حملت الاسم فقط ؛ ونقصد بذلك أن نظام استرجاع المعلومات كمفهوم قد ارتبط أكثر بالبيئة الرقمية، على الرغم من أن مختلف مرافق المعلومات التقليدية ذات المبنى والجدران هي نظم لاسترجاع المعلومات.

يختلف حال المستفيد في التعامل مع الوثائق بين النظم التقليدية والنظم الرقمية، وتؤدي الوثائق ذاتها دورًا مهمًا في التفاعل بين النظام والمستفيد ؛ حيث تتيح النظم التقليدية دائمًا المصادر الأولية أمام المستفيد دون إتاحة الإشارات البيليوجرافية فقط كحال بعض النظم الرقمية. فالمكتبة كشكل من أشكال نظم استرجاع المعلومات التقليدية لا تضع في فهارسها إشارات ببليوجرافية لوثائق ليست بحوزتها، فكما هو معروف أن لكل بطاقة في فهرس المكتبة كتابًا على أرفف هذه المكتبة. أما النظم الرقمية فقد عملت في مراحلها الأولى على احتواء الإشارات الببليوجرافية التي قد تصاحب في بعض الأحيان مستخلصات للوثائق، وقد دعمت القدرات المتطورة لأنظمة تخزين المعلومات إمكانات احتوائها الآن على النص الكامل، وليس هذا

فقط، بل إنه قد أصبح لديها القدرة على البحث والاسترجاع باستخدام كل الكلمات الواردة في متن النصوص الكاملة.

ثانيًا : الوثائق الرقمية:

ترتبط الوثيقة الرقمية دائمًا بالاستخدام والإتاحة من خلال الحاسبات الآلية والشبكات. ومن الممكن أن تشكل الوثائق الرقمية ذات مستوى محدد من المعلومات نظامًا فرعيًا من نظام أكبر تقليدي ؛ ومثالاً لذلك التسجيلات الببليوجرافية التي توجد داخل الفهرس الآلي لأي مكتبة ، إنما هي وثائق داخل نظام استرجاع (الفهرس الآلي) الذي بدوره يرتبط بنظام استرجاع أكبر هو مرفق المعلومات. وفي حال دخول الوثائق إلى مستودع المعلومات، فإنها جميعًا يطلق عليها وثائق رقمية ، ويأتي اختلافها فقط في أصل هذه الوثائق التي اعتمدت على أصل تقليدي ثم تحولت إلى الرقمية ، أم أنها نشأت بشكل رقمي مباشر.

١- مواد رقمية لها أصل تقليدي :

تتمثل هذه الوثائق في مختلف مواد المعلومات التي تتحول إلى الرقمية مع رغبة النظام العام في التحول إلى الشكل الآلي. ومن أكثر الأمثلة وضوحًا على هذه الوثائق، ما يحدث لبطاقات الفهارس التقليدية عند اعتماد قرار مرفق المعلومات بتحويل العمليات إلى الشكل الرقمي.

٢- مواد رقمية خالصة:

وهذه المواد هي تلك المنتجة والمصممة من خلال الحاسب الآلي وليس لها أصل في البيئة التقليدية ، ومن نماذجها صفحات ومواقع الويب، والرسوم

الحيوية animations ، هذا إلى جانب وثائق البيئة الإدارية المنتجة بواسطة الحاسب الآلي.

٢ / ١ / ١ / ٣ من حيث درجة المعلومات:

تختلف نظم استرجاع المعلومات فيما بينها في درجة تلبية حاجات المستفيدين من المعلومات، ويأتي هذا التفاوت ليس لقصور في عمليات النظام أو المعالجة وإنما من طبيعة المواد المعلوماتية التي يحتويها النظام. ولقد أحدثت الويب تغيرًا جذريًا في تلبية حاجات المستفيدين للمعلومات، حيث مكنت قدرات الويب في الرقمنة digitization والتخزين من انتشار نظم استرجاع النصوص الكاملة التي تعد الغاية الكبرى لمختلف المستفيدين في الحصول على المعلومات. ولذلك فهي تقسم إلى ثلاثة أنواع:

أولاً : وثائق أولية:

لقد أتت صفة الأولية هنا من طبيعة هذه الوثائق في كونها أول ما نشر من معلومات في صيغتها الكاملة وما تحمله من معلومات هي الأولى في نشر الفكرة أو الأفكار. ومن أمثلة هذه الوثائق مقالات الدوريات والكتب والرسائل الأكاديمية... وغيرها. ويمكن لمستخدم الويب الآن استرجاع مختلف النصوص الكاملة في أشكال مختلفة من الملفات مثل pdf , doc، والجانب الأكثر أهمية من ذلك هو التعامل مع النصوص الكاملة ليس فقط في الاسترجاع والبحث وإنما، في المعالجة الفنية والتنظيم من خلال جعل كل الكلمات الواردة في الوثيقة مداخل لاسترجاع هذه الوثيقة.

ثانيًا : وثائق من الدرجة الثانية :

تتمثل الوثائق من الدرجة الثانية في تلك التسجيلات أو المستخلصات أو الكشافات التي تسعى إلى توثيق المعلومات الأولية. وتسمى بالدرجة الثانية لكونها تعطي فقط معلومات عن المعلومات الأولية التي جاءت في النوع الأول. وأهم نماذج نظم استرجاع المعلومات التي تعتمد على الوثائق من الدرجة الثانية هي قواعد البيانات الببليوجرافية أو قواعد كشافات الرسائل الجامعية أو كشافات الدوريات.

ثالثًا : وثائق من الدرجة الثالثة :

يختلف هذا النوع عن سابقيه، فلم توجد أي من نظم استرجاع المعلومات سواء التقليدية أو الرقمية التي احتوت هذا النوع، غير أن الويب قد قدمت لنا نموذجًا من نظم استرجاع المعلومات الذي يشكل استرجاع المعلومات من الدرجة الثالثة في شكل معلومات أولية أمام مستخدمي الويب، وهذا النوع هو محركات بحث قواعد البيانات، حيث يستطيع محرك البحث الدخول إلى عمق قواعد البيانات واسترجاع ما يلائم استراتيجية البحث التي صاغها المستفيد وأدخلها إلى واجهة البحث العامة.

يحمل نظام الاسترجاع العالمي للمعلومات (الويب) مجموعات غير متناهية من أنواع الملفات التي تحمل وسائط مختلفة من المعلومات. فإن جاز القول تعد الويب بمثابة نظام استرجاع تتشابك عناصره من خلال بروتوكول الإنترنت، ويتمثل مستودع معلوماته في تجمع مساحات التخزين لكل خادمات الويب المشتركة في الشبكة، وتقوم عمليات المعالجة داخله لمختلف أنواع الوثائق

التي تحتويها. وقد حرص الكثير من الدراسات العالمية والعربية التي لا يمكن حصرها على الاتجاه إلى رأس الويب ومحركات البحث والمعالجة الفنية لمجموعات الوثائق التي تحملها خادمات الويب. وأصبح الآن لكل وسيط أو شكل من أشكال المعلومات المعايير والتقييس الخاص به.

٢ / ١ / ٢ اللغة:

اللغة نظام من الإشارات التي تعبر عن الأفكار وفي مجال المكتبات تستخدم الإشارات الرقمية والحرفية ، والنصوص المسجلة التي تعبر عن المفاهيم التي يمكن معالجتها في عمليات التحليل الموضوعي و الاسترجاع، والنص كيان معرفي له دلالات إيحائية أو مفاهيم قابلة للتفسير حسب وجهات النظر للمتلقي أو المجال الموضوعي أو مناهج أو أغراض الدراسات المختلفة، وتتفاوت لغة النص حسب وظيفة أو أسلوب الكتابة الذي يؤدي دورًا مهمًا في تشكيل النص أو المعلومة. ويمكن تصنيف النص من حيث البناء اللغوي إلى ثلاث فئات من الكلمات حسب الصفات الدلالية وهي[١]:

١. **المفهوم الدال** : وهي المصطلحات أو الكلمات التي تحمل دلالات موضوعية مستقرة ومتفقًا عليها بتسمية الأفكار أو الأشياء .

٢. **المفهوم غير الدال** : و هي المعاني العامة أو النسبية المتغيرة التي تلازم المفردات ولا يمكن تحديد دلالتها الموضوعية بشكل ثابت .

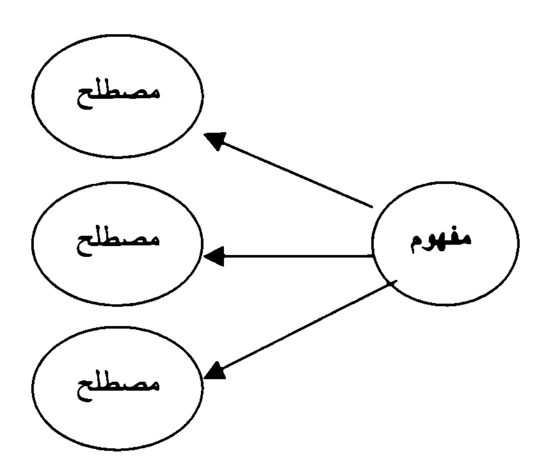
٣. **الحروف و الأدوات اللغوية** : اللازمة لتكوين النص باللغة الطبيعية .

(١) هانئ محيى الدين عطية. **محاضرات غير منشورة.**

وفي نظم الاسترجاع يتم التعامل مع المفهوم الدال، أما المفهوم غير الدال ، الحروف والأدوات فإنها توضع في قائمة الاستبعاد .

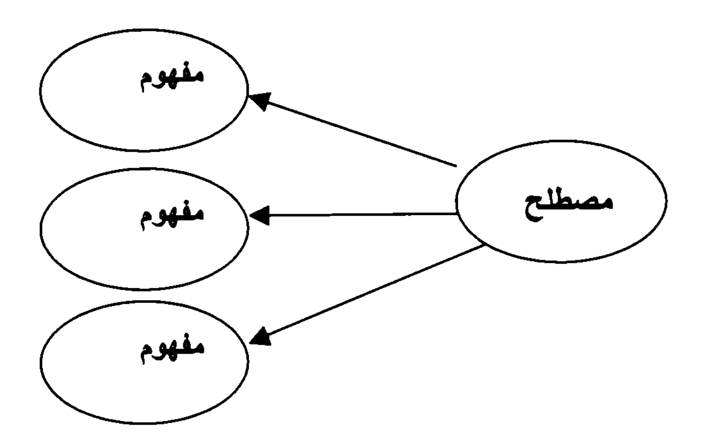
ويعد المفهوم الدال في نظم الاسترجاع هو الأساس الذي يجمع بين المستفيدين والنظام ولكن هناك الكثير من المشكلات التي تصاحب المفهوم الدال في نظم الاسترجاع في حالة استخدام اللغة الطبيعية وهي :

✓ **الترادف** :- ويعني وجود أكثر من مصطلح واحد للدلالة على موضوع واحد أو مفهوم معين.



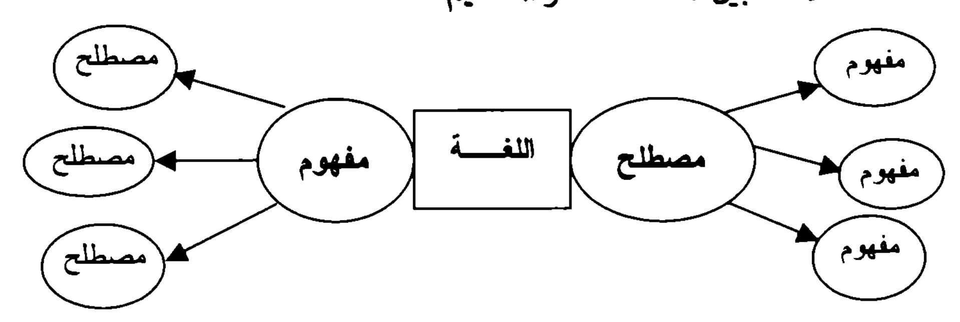
الشكل رقم (٢ -١٣ أ) الترادف بين المصطلحات في اللغة

✓ **الاشتراك اللفظي** :- ويعني وجود أكثر من مفهوم واحد للدلالة على مصطلح معين .



الشكل رقم (٢ -٣ ب) الاشتراك اللفظي بين المصطلحات في اللغة

✓ **العلاقات بين المصطلحات و المفاهيم :**



الشكل رقم (٢ -٣ ج) العلاقة بين المصطلحات والمفاهيم في اللغة

وهي وجود مصطلحات تمثل مفاهيم معينة في البيئة التي يجري فيها البحث أو ذات الفرع من العلم المتخصص .

وتحاول نظم الاسترجاع التغلب على هذا النوع من المشكلات باستخدام اللغة المقيدة وذلك من خلال:

أ- التحكم في المترادفات والاشتراك اللفظي باستخدام قوائم رؤوس الموضوعات أو المكانز .

ب- التحكم في العلاقات بين المصطلحات أو المفاهيم رأسيًا أو أفقيًا من خلال شبكة الإحالات.

فأهداف اللغة المقيدة هي تحقيق تطابق أو تكافؤ تام بين المصطلحات اللغوية والأشياء أو الأفكار أو المفاهيم التي تدل عليها وتحديد ما بين المصطلحات أو المفاهيم من علاقات أفقية ورأسية .

ولقد أدت النتائج والاختبارات الخاصة بنظم الاسترجاع التخلي عن اللغات المقيدة لصالح اللغة الطبيعية وذلك لعدة أسباب :

١- تزايد عدد الوثائق التي تحتاج إلى تكشيف .

٢- تزايد كم نصوص اللغة الطبيعية التي يمكن البحث من خلالها وخاصة النصوص الكاملة للوثائق.

٣- تكلفة وضع وصيانة اللغة المقيدة .

٤- تأخر ظهور بيانات الوثائق في المصادر الوراقية نتيجة الوقت اللازم لتكشيفها .

٥- تكلفة اللغة المقيدة وبعدها عن اللغة الطبيعية .

وعمومًا فإن هناك اتجاهًا سائدًا الآن للتوفيق بين اللغات المقيدة واللغات الطبيعية؛ لأنه يجمع بين مزايا كلٍّ منهما .

٢ / ١ / ٢ / ١ لغات المعالجة والبحث:

تستخدم نظم استرجاع المعلومات كلاً من اللغة الطبيعية الحرة واللغة المقيدة المضبوطة ليس فقط في مرحلة بحث واسترجاع المعلومات، وإنما تستخدمهما أيضًا في مرحلة المعالجة ؛ ويمثل ذلك الاعتماد على طريقتي

التكشيف بالتعيين والتكشيف بالاقتباس. واستخدام نظام استرجاع المعلومات للتكشيف بالاقتباس وإخراج الكلمات الدالة من النص الكامل هو جوهر اللغة الطبيعية الحرة، أما استخدام نظام الاسترجاع للتكشيف بالتعيين واستخدام قوائم رؤوس الموضوعات والمكانز هو أساس اللغة الطبيعية. ومن ثم فإن نظام استرجاع المعلومات يرتبط باللغة المستخدمة في مرحلة المعالجة الفنية لتكون هي ذاتها المستخدمة في عمليات البحث والاسترجاع، وهذا إنما يأتي من تحقيق التوافق بين الكلمات المفتاحية الدالة على الموضوع والكلمات المفتاحية المكونة لاستراتيجية البحث.

إن دراسات التطوير للغة داخل نظم استرجاع المعلومات لم تقتصر على التقدم في معالجة اللغة الطبيعية أو المقيدة، وإنما امتد الأمر إلى تطوير قدرات نظم استرجاع المعلومات على معالجة اللغات الحية المستخدمة في كتابة الوثائق والإنتاج الفكري على مستوى العالم. ومن هذه الدراسات ما تم عن طريق naserdine semmar وزملائه في تطوير قدرات نظم استرجاع المعلومات للعمل على تنظيم واسترجاع المعلومات المسجلة بلغات مختلفة. من خلال توفير برامج للتحليل اللغوي والإحصائي داخل نظام الاسترجاع يأتي دورها في العمل عند إضافة الوثائق ومعالجتها. ويعمل هذا النظام اعتمادًا على مجموعة من برامج تحليل اللغة سواء المستخدمة في استفسار البحث أم تلك المستخدمة في كتابة الوثائق. وعلى ذلك فإن ثمة ترابطًا قد أحدثته هذه الدراسة بالجمع بين آليات المعالجة والبحث مثل المنطق البوليني وبرامج تحليل اللغة. من البرامج المستخدمة في هذا النظام ما يلي(١):

---

(١) Semmar, naserdine. A cross language information retrieval system, laboratory of multilanguage multimedia, 2005, cited 11/11/2007, cited at wwwlist.cea.fr/.../gb/Alger2005_semmar_a_cross_language_information_retrieval_system.pdf .

١- **برامج التحليل اللغوي** : linguistic analyzer وهي التي تقوم بالمعالجة اللغوية لكل من استفسار البحث والوثائق المكشفة.

٢- **برامج التحليل الإحصائي** : statistic analyzer وهي تقوم فقط بتكشيف الوثائق.

٣- **برامج تهذيب البحث** : reformulator وتختص بتوسيع سؤال البحث وإجراء البحث داخل النتائج المسترجعة.

٤- **برامج الربط أو المضاهاة** : comprator وهي تقوم بتحديد التشابه الدلالي بين سؤال البحث والوثائق المكشفة.

٥- **برنامج المكشف** : indexer وهو الذي يقوم بتخزين الوثائق داخل قاعدة البيانات.

٦- **آليات أو محرك البحث** : search engunes وهو الذي يقوم بالبحث داخل الكشاف والوصول إلى أقرب الوثائق لاستفسار البحث، ثم دمجها معًا في شكل قائمة واحدة مرتبة.

ولكي يستطيع برنامج التحليل اللغوي التعامل مع الوثائق المكتوبة بلغات مختلفة، فقد خصص البرنامج مجموعة من أدوات العمل والمعالجة لكل لغة على حدة، منها:

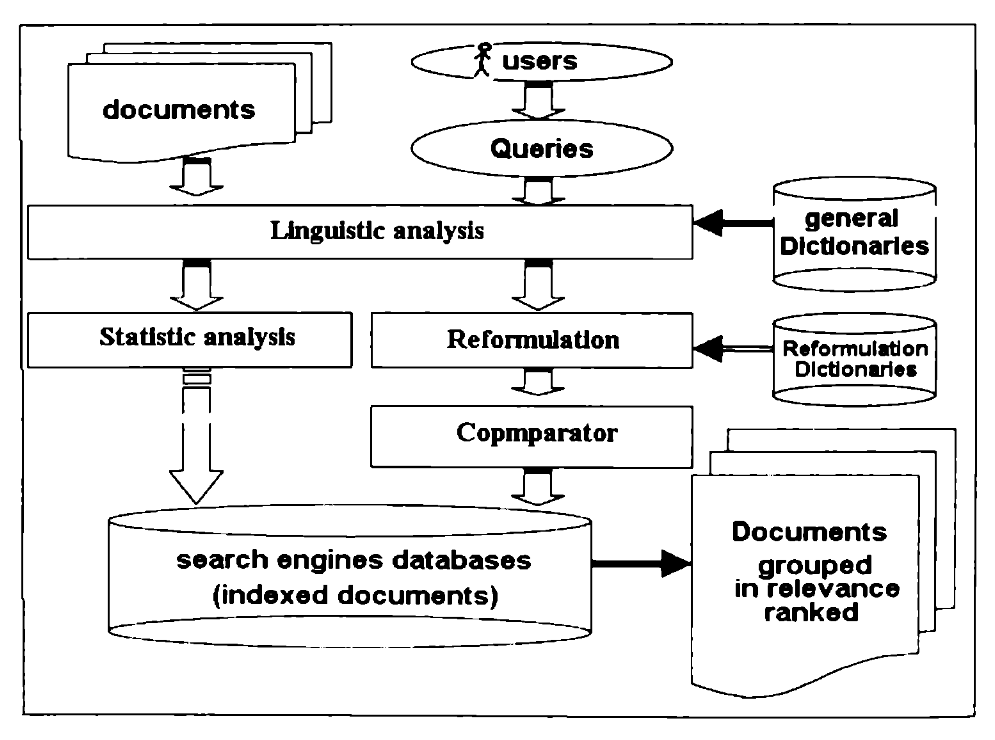
أ- **معجم لغوي متكامل** يحمل الكلمات بأشكالها المختلفة، بالإضافة إلى احتوائه على نماذج الحروف والأرقام والإشارات اللغوية المختلفة.

ب- **معجم أحادي اللغة** يستطيع معالجة الاستفسارات القادمة إلى النظام وتحليلها ومعالجتها من حيث المقابلات والمعنى والتراكيب لكلمات السؤال.

ج- **معاجم متعددة اللغات** للربط بين اللغات المختلفة والترجمة من لغة إلى أخرى.

د- **مجموعة من القواعد** لتحليل الكلمات وردها إلى الأصل.

ويمثل الشكل رقم (٢-٤) نظام استرجاع المعلومات متعدد اللغات، مع بيان آليات العمل والتحليل اللغوي للوثائق داخل نظام استرجاع المعلومات:



الشكل رقم (٢ -٤) عمل برامج التحليل اللغوي في نظم استرجاع المعلومات متعدد اللغات

ومن حيث المعالجة والعمليات الفنية؛ فإنه يمكن تناول اللغة داخل نظام الاسترجاع في شقيها الأساسيين وهما:

أولاً : اللغة الطبيعية:

وتسمى أيضاً اللغة غير المقيدة، أو لغة النص الحر، فهي تستخدم الكلمات الحقيقية للوثيقة كألفاظ تكشيف، ويطلق على هذا النوع من التكشيف المشتق؛ لأن المصطلحات المستخدمة في الكشاف تشتق مباشرة من نص الوثيقة أو من عنوانها أو مستخلصها. وهذا النوع من التكشيف لا تكون فيه جهود مبذولة للسيطرة على العلاقات الدلالية بين مصطلحات التكشيف أو تقييده، أي ما قد يكون تم وصفه ثم تكشيفه دون أن تكون هناك روابط بين تلك المصطلحات المتاحة للمستفيد بنظام الاسترجاع، وتشتمل اللغة غير المقيدة على المترادفات، وأشباه المترادفات والأشكال واللهجات المختلفة للكلمة .

مزايا البحث باللغة الطبيعية الحرة :

١- توافر أوجه متعددة للبحث وبدرجة عالية من الشمول والتخصص، فكل المفردات التي استعملها المؤلف يمكن أن تتحول إلى مصطلحات كشفية، سواء أكانت أسماء أشخاص أم هيئات أو منظمات أو مواقع ...الخ.

٢- قدرة النظام على استيعاب التطورات والتحديثات المستمرة للمفردات والمصطلحات الموضوعية بسبب دخولها المباشر من نص الوثيقة الذي يحرص المؤلفون من خلاله على مواكبة المستجدات العلمية والموضوعية .

٣- بناء الكشاف من مفردات التسجيلات المقروءة آليًا أقل كلفة وأسهل في عملية الإعداد والتحديث .

٤- يقدم مرونة أكبر في إجراء البحث بالربط اللاحق بين المفردات دون أن يدمر الهوية المستقلة لكل مفردة .

٥- هذا النظام هو الأنسب في عمليات البحث التي يقوم بها المستفيد بشكل مباشر أو من خلال وسيط متخصص بالموضوع .

عيوب البحث باللغة الطبيعية الحرة:

وعلى الرغم من الإيجابيات السابقة التي تتسم بها اللغة الطبيعية، فإنها تحمل على الجانب الآخر بعض السلبيات ، منها الآتي :

١- الاشتراك اللفظي . حيث قد يتحمل اللفظ الواحد أكثر من معنى.

٢- كثرة المترادفات . وتؤثر المترادفات سلبًا على عمليات البحث في زيادة عامل الفقد، فقد يستخدم المستفيد في استراتيجية البحث كلمات لم يستخدمها المؤلف للدلالة على المعنى نفسه.

٣- اختلاف المعنى والإملاء بين الكلمات والألفاظ المتحدة في البيئات المختلفة.

ثانيًا : اللغة المقيدة:

وهي اللغة التي تتحدد مفرداتها وتحسم مشكلاتها الدلالية وتستقر قواعدها النحوية من البداية بحيث تكون بين يدي المكشفين أدوات جاهزة مشتملة على المداخل الكشفية في أشكال محددة ينبغي التقيد بها في التعبير عن ناتج التحقيق من المحتوى الموضعي لأوعية المعلومات وهي تتمثل في خطط التصنيف الحصرية لتصنيف ديوي العشري وتصنيف مكتبة الكونجرس

والتصنيف العشري العالمي، كما تتمثل أيضاً في خطط التصنيف الوجهية أو التحليلية التركيبية التي وضع أسسها العالم "رانجاناثان" وتتمثل أيضاً في قوائم ورؤوس الموضوعات المعيارية أو المقننة أو الملفات الاستنادية الموضوعية، كذلك تتمثل في المكانز على اختلاف مستوياتها وتنوع أشكالها.

إيجابيات اللغة المقيدة :

تتسم اللغة المقيدة ببعض الإيجابيات منها التالي:

١- تزيد من كفاءة الاطراد بين كل من المؤلف في التعبير عن المعنى في كلمة واستخدام نفس الكلمات كمداخل كشفية للتعبير عن موضوع ومحتوى الوثيقة.

٢- تساهم أدوات اللغة المقيدة مثل المكانز ورؤوس الموضوعات في خلق الاطراد بين المكشفين المختلفين في الأزمنة المختلفة وفي الأماكن المختلفة.

٣- تساهم أدوات المكانز ورؤوس الموضوعات في توضيح الروابط المنطقية بين المداخل الكشفية، من خلال استخدام شبكة الإحالات وعلى ذلك يستطيع المستفيد التعرف إلى المصطلحات المرتبطة في موضوع واحد أو في موضوعات مترابطة.

٤- تزيد اللغة المقيدة من نسبة التحقيق في النتائج المسترجعة.

وعلى الجانب الآخر، فإن ثمة سلبيات تحدث عنها طلال الزهيري تكمن في[1]:

---

(١) طلال ناظم الزهيري. **العوامل المتصلة بنظام استرجاع المعلومات وتأثيرها على كفاءة الاسترجاع**، تاريخ الاستشهاد ٢٠٠٧/١٢/١٠، مسار المصدر
http://azuhairi.jeeran.com/files/81729.doc

١- تحتاج المكانز ورؤوس الموضوعات دائمًا إلى عمليات التحديث وإضافة المصطلحات الجديدة للتعبير عما يستجد من نقاط موضوعية في المجال.

٢- أن أي محاولة تفرض على المستفيد في استخدام اللغة المقيدة في صياغة استفساره قد تؤدي إلى استقراره على أقل مما يريده فعلاً .

٣- التغطية الشاملة للنتائج التي توفرها اللغة المقيدة، غالبًا ما تكون على حساب مستوى التخصص، بعبارة أخرى، فإن الاستدعاء العالي يكون دائما على حساب الدقة .

٤- التكاليف العالية لبناء وتطوير أدوات ضبط المصطلحات من مكانز وقوائم رؤوس موضوعات. وبسبب هذه التكاليف يصعب في كثير من الأحيان تحديث هذه الأدوات لتكون مواكبة للتطورات الجارية في المجالات العلمية المختلفة، مما يؤدي إلى وجود ضعف في تغطيتها للمصطلحات المستحدثة.

٥- نظم البحث باللغة المقيدة تحد من إمكانية المستفيد في إجراء العملية البحثية دون وساطة، وتدخل الوسيط غالبًا ما يترتب عليه زيادة في تكلفة تنفيذ هذه العمليات.

٦- إن أي خطأ يرتكبه المكشف في اختيار المصطلح الموضوعي المعبر عن المحتوى الموضوعي للوثائق ، يؤدي إلى خسارة محتملة عند الاسترجاع لتسجيلات يمكن أن تكون ذات صلة بموضوع البحث .

٢ / ١ / ٢ / ٢ أساليب البحث:

٢ / ١ / ٢ / ٢ / ١ البحث بالتصفح:

يأخذ البحث بالتقسيم الموضوعي أكثر من مسمى، حيث يطلق عليه أيضًا اسم التصفح. كما أن تطبيق هذا البحث يختلف من مجموعة إلى أخرى، فبينما تقوم بعض نظم الاسترجاع بالاعتماد على أحد أنظمة التكشيف أو التصنيف في عمل تقسيم موضوعي من الأقسام العامة إلى الأقسام الأكثر تخصصيًا، تقوم بعض النظم الأخرى بعمل أشكال للتقسيمات الموضوعية الخاصة بها. ويعتمد أسلوب التصفح على شكلين أساسيين من عرض محتويات نظم استرجاع المعلومات ؛ أولهما : عرض كل مجموعات الوثائق التي توجد داخل قاعدة البيانات، مع تقسيم هذه المجموعات في أقسام عامة تحوي مجوعات أكثر تخصصًا، وتكون هنا الوحدات في شكل تسجيلات ببليوجرافية بسيطة تصحبها بعض اللقطات في حالة الوسائط المتعددة، على أن يصاحب هذه اللقطات بعض الحقول النصية التي توضح ماهية هذه الوثيقة وموضوعها؛ ليتمكن المستفيد من الحكم على صلة هذه الوثيقة بموضوع بحثه. ويتناسب هذا الأسلوب في التصفح فقط مع قواعد البيانات ذات الحجم الصغير من نظم الاسترجاع، لأن قواعد بيانات النظم كبيرة الحجم لن يجد المستفيد الوقت الكافي للتعرف إلى كل ما فيها وتحديد ما يناسب احتياجاته الموضوعية. أما الشكل الثاني فهو تكشيف كل مجموعات الوثائق داخل قاعدة البيانات ثم وضع المصطلحات في الترتيب المنطقي لها، على أن تكون في ترتيب من الأعم إلى الخاص إلى الأكثر تخصصيًا، وربط كل مجموعة من الوثائق بالمصطلح الذي يعبر عن موضوع

هـذه الوثائق باستخدام تقنية الربط الفائق. وهذا من الممكن أن يكون ملائمًا لطبيعة قواعد بيانات نظم الاسترجاع على الويب التي يزداد حجمها بشكل سريع ومتلاحق.

علاقة التصفح باللغة المضبوطة:

يعتمد التكشيف على اختيار الكلمات الدالة التي تصف محتوى وسيط المعلومات، حيث يعد إحدى عمليات التحليل الموضوعي التي تقف عند مستوى المحتوى الموضوعي والتعبير عنه بالكلمات أو المصطلحات الدالة، وهو خلاف ما يحدث في عمليات التحليل الأخرى مثل التصنيف في التعبير عن الموضوع بواسطة رموز من حروف أو أرقام، وعملية الاستخلاص التي تتعدى مستوى الكلمات إلى مستوى جمل واصفة مكونة شكل المستخلص النهائي. وبرز مصطلح لغة التكشيف من خلال طبيعة المصطلحات المستخدمة للتعبير عن المحتوى الموضوعي ؛ فإذا ما تم الاعتماد على قائمة من المصطلحات المعدة والمختارة مسبقًا، أصبحت اللغة مقيدة أو محكومة أو مضبوطة، أما إذا تم الاعتماد على طبيعة الكلمات التي وردت في النص أو العنوان وفقًا لما اختاره المؤلف، أصبحت اللغة طبيعية. ويتم الاعتماد فقط في هذه الحالة على طبيعة آليات البحث داخل قاعدة البيانات وليس على المطابقة بين قائمة المصطلحات والمداخل التي وضعت أسفلها مصادر المعلومات داخل قاعدة البيانات[1].

ويمثل التصفح برؤوس الموضوعات أو المصطلحات داخل الويب نموذجًا من نماذج الاعتماد على اللغة المقيدة في التكشيف وتحليل محتوى مصادر الويب،

(١) سيد ربيع سيد إبراهيم. **محركات بحث الصور الثابتة على الإنترنت : دراسة تحليلية.** ٠ الرياض : مكتبة الملك فهد الوطنية، ٢٠٠٧.

ذلك لما يرتبط به من تقيد مستخدم نظام الاسترجاع بقائمة مصطلحات هرمية التنظيم يتم ربطها بمجموعة مصادر المعلومات التي تتوافق مع دلالة كل من المصطلحات. وهناك من أدوات بحث الويب مثل Yahoo ما يعتمد على قائمة من المصطلحات الخاصة بالموقع تندرج أسفل منها صفحات الويب التي تلبي الحاجة الموضوعية الخاصة بكل من المصطلحات. ويمكن في هذه الحالة الاعتماد على أي من المكانز الخاصة بالتكشيف في مجال محدد أو مجالات متعددة بصياغة شكل من أشكال التصفح الداعم للتحكم في مجموعة ثابتة من المصطلحات تتكون على أثرها قوائم رئيسة ثم فرعية تدرج أسفل منها محتويات قاعدة البيانات ككل. وقد اعتمدت بعض أدوات البحث على نظم تصنيف معدة مسبقًا مثل تصنيف ديوي والتصنيف العشري العالمي، أو أن تعتمد بعض المواقع الأخرى على صياغة أشكال من نظم التصنيف تلائم المحتويات الخاصة بها من مصادر الويب. وقد ذكر هانئ عطية بعضًا من جهود تصنيف المعلومات على الويب في سرد مجموعة من نظم التصنيف المعدة مسبقًا أو ذاتية البناء، بالإضافة التي تقسيم تلك النظم إلى أنماط عامة متعددة مثل نظم التصنيف المكتبية كنظام مكتبة الكونجرس ، ونظم التصنيف القومية كالتصنيف الأساسي في هولندا والتصنيف القومي السويدي للمكتبات العامة، ونظم التصنيف الموضوعية كالتصنيف الخاص في مجال الحاسب الآلي والتصنيف الموضوعي للرياضيات، ونظم التصنيف المولدة في البيئة الرقمية كتصنيف بوابة Edinburgh Engineering Virtual Library وتصنيف بوابة Social Science Information Gateway. وقد أكد الكاتب على أنه توجد على الأقل ١٦ تصنيفًا تخدم تنظيم المعلومات على الويب ثلاثة منها عالمية، واثنان قوميان، وأربعة موضوعية، وسبعة خاصة، مما يعني

البحث الدؤوب حول نظام تصنيف يوفر الدرجة الأفضل من التحقيق داخل النتائج المسترجعة. ومن بين ما أكدت عليه هذه الدراسة أيضًا أن أغلب المواقع قد لجأت إلى استخدام نظم التصنيف العالمية خلاف النظم القومية، وذلك لما لها من ثبات وتجربة سابقة في الاستخدام لاسترجاع المعلومات. كما جاء أيضًا من بين النتائج أن أغلب المواقع قد لجأ إلى استخدام المكانز بدلاً من قوائم رؤوس الموضوعات لما لها من سمات مهمة مثل الاعتماد على المصطلحات المفردة؛ إضافة إلى التخصص الموضوعي في مجالات محددة يغطيها مجال موقع الويب فقط[1].

مميزات أسلوب التصفح :

١- يتميز أسلوب التصفح عن أسلوب الكلمات المفتاحية بتحقيق التفاعلية بين النظام والمستفيد؛ حيث يجد المستفيد في بحثه خلال قوائم المصطلحات إشارات ترشده إلى الوقوف على ما يريده من معلومات و مساعدته في الوصول إليها .

٢- يعمل أسلوب التصفح أكثر على تصحيح مسار المستفيد في الوصول إلى المعلومات ، وذلك من خلال تجهيز و إعداد المصطلحات أمام المستفيد دون الحاجة إلى محاولات الصواب و الخطأ في صياغة كلمات البحث .

٣- لا يحتاج أسلوب التصفح الجهد اللازم لمعرفة معاملات وآليات البحث داخل محركات البحث التي تختلف من محرك إلى آخر ؛ حيث يوفر

(١) هاني محيى الدين عطية. **جهود تصنيف المعرفة في البيئة الإلكترونية : دراسة استكشافية** .....

التصفح مصطلحات البحث بدون الحاجة إلى معاملات البحث للربط بينها و صياغة إستراتيجيات البحث .

٤- يوفر التصفح المرئي لوحدات الصور الرقمية سهولة الاختيار والبحث للمستفيد من خلال التعرف إلى موضوعات الصور بطريقة أقل من تلك المستخدمة مع قوائم المصطلحات.

٥- يتميز أسلوب التصفح بالمصطلحات بأن المستفيد يستطيع من خلاله الإبحار داخل المصطلحات ذات الدلالة على محتوى و موضوع وثائق نظام الاسترجاع، وهو ما يوفر التعرف إلى السمات الموضوعية للوحدات داخل قاعدة البيانات .

٦- تستطيع المستويات المختلفة من المستفيدين التعامل بكفاية كبيرة مع أسلوب التصفح ؛ حيث إن أسلوب البحث بالكلمات المفتاحية موقوف فقط على المستفيدين أصحاب المهارات معاملات وآليات البحث على الويب .

٧- يؤدي استخدام أدوات اللغة المقيدة في التصفح المكانز وقوائم رؤوس الموضوعات إلى تحقيق درجات أكبر من التحقيق وخفض معدل الاستدعاء داخل نظم استرجاع المعلومات، وهو ما يعتمد على الدقة في التكشيف والبحث.

عيوب أسلوب التصفح :

١- إن استخدام أسلوب التصفح بالمصطلحات يتطلب القدر الكبير من الدقة في تحليل محتوى الوثائق التي تحمل معاني متباينة ومتداخلة مثل

الصور الرقمية، وقد يؤدي تبني هذا الأسلوب إلى تشتت المستفيد في حالة الضعف الدلالي للمصطلحات .

٢- إن أسلوب التصفح المرئي لوحدات الصور يعمل بكفاية أكثر داخل قواعد البيانات الصغيرة، وهو مالا يصلح للعمل داخل قواعد البيانات كبيرة الحجم ؛ حيث يحتاج المستفيد الكثير من الجهد والوقت لتصفح مجموعات الصور.

٣- يحتاج بناء قوائم المصطلحات الدالة داخل محركات البحث إلى إشراف العنصر البشري وهو يخالف طبيعة محركات البحث التي تعتمد على برامج ذاتية العمل قد لا ترقى إلى اعتماد المصطلحات ذات الدلالة لمحرك البحث .

٢ / ١ / ٢ / ٢ / ٢ البحث الحر:

يحتاج أسلوب الكلمات المفتاحية في بحث النصوص الكاملة بلغتها الطبيعية إلى مجموعة من آليات البحث القادرة على صياغة مختلف استراتيجيات البحث بما يحقق الدرجة المرجوة من التحقيق في استرجاع الوثائق الرقمية. وتعمل الآليات العامة مثل معاملات الربط البوليني ومعاملات البحث الحر مع مختلف لغات البحث، كما تعمل آليات بحث مثل البحث بالتقابل، والبحث بالترجمة، والبحث بالمعنى، والبحث بالمشترك اللفظي... وغيرها لمعالجة خصائص اللغة العربية. ويمكن القول إن الاستخدام الأمثل لآليات بحث النظام يمكن معه تحقيق أفضل استدعاء مع أفضل تحقيق لوحدات النظام المستدعاة.

ويتميز أسلوب الكلمات المفتاحية بالكثير من المميزات التي يمكن تحقيقها مع مجموعة من العيوب التي يمكن تحويلها إلى جوانب إيجابية، قد حثت عليها دراسة[1] لإستراتيجيات بحث الويب، وهي مثل :

فمن مميزات البحث بالكلمات المفتاحية:

١- يمكن للمستفيد القفز السريع إلى الصفحات المحددة لخدمة الموضوع إذا ما أتقن استخدام آليات البحث.

٢- يستطيع المستفيد الحصول على صفحات الويب غير المدرجة في قوائم البحث للأدلة إما لعدم وجود المهارة في معرفة الموضوع أو لعدم وجود الوقت الكافي للتعامل معها.

٣- يرتفع معدل حداثة صفحات الويب داخل أدوات البحث المعتمدة على برامج الزاحف والعنكبوت نظرًا لتعاملها السريع في الإضافة والمعالجة المعلوماتية أسرع من الإنسان.

وبالمقابل، فإن من عيوب البحث بالكلمات المفتاحية:

١- قد لا يستطيع المستفيد الحصول على صفحات الويب الملائمة إذا ما افتقر إلى الكلمات المفتاحية عن الموضوع.

٢- لأن الكلمات المفتاحية هي العمود الفقري لمحركات البحث، فإنه لا يمكن بذلك تصفح محتويات أي من المحركات إذا ما افتقر الباحث

(١) Keyword Searches.[with out publisher] , 2002 , cited 10/10/2002 , cited at : http://www.concentric.net/~Rkriesel/Search/S-Keywords.htm.

إلى كلمات إرشادية قد يحصل عليها من خلال البناء الهرمي للموضوعات في التصفح الموضوعي.

٣- البحث بالكلمات المفتاحية داخل محركات البحث يستدعي معه الغث والثمين من المعلومات خلاف أدلة البحث التي تخضع للتقييم البشري.

٤- لا يضم محرك بحث بعينه كل مجموعات الويب، مما يستلزم معه بحث أكثر من محرك.

٥- يحتاج البحث بالكلمات المفتاحية إلى مهارة التعرف إلى آليات البحث المساعدة للكلمات المطروحة على نظام الاسترجاع.

٢ / ١ / ٢ / ٣ استراتيجيات البحث (صياغة اللغة):

٢ / ١ / ٢ / ٣ / ١ المنطق البولياني : bolean logoic

وضع الرياضي البريطاني George Boole في عام ١٨٥٠م العوامل البوليانية التي نسبت إليه، وقد وضعها اعتماداً على قواعد الجبر والمنطق. وتتمثل عوامل المنطق البولياني الرئيسة في ثلاثة عوامل هي: and , or , not (و ، أو، ماعدا)، ويؤدي كل منها وظيفة مختلفة عن الآخر.

ويتم تمثيل كل عامل من تلك العوامل بشكل من أشكال فن Venn diagrams، التي تسمح للباحثين بالتخطيط وإيضاح إستراتيجية البحث في الحالات التي يتم فيها استخدام أكثر من مصطلح لصياغة الإستراتيجية(١).

---

(١) Whitson , Donna L & Amstutz , Donna D. Accessing Information in a Technological age .- Malabar, Florida : Krieger publishing company , 1997 .- p68.

وفيما يلي نوضح تلك العوامل مع أشكال فن الخاصة بكل منها:

أ. and (و) : عامل الضم أو الاقتران conjunction ، وباستخدام هذا العامل يتم فقط استرجاع التسجيلات التي تضم كل أوجه الاستفسار، بمعنى أنه يجب أن تقترن في التسجيلات المسترجعة جميع مصطلحات البحث المستخدمة، فعلى سبيل المثال لو تم إجراء بحث حول **الكتب والدوريات**.

ب. or (أو) : للفصل disjunction ، ويسترجع التسجيلات التي يظهر فيها أي مصطلح من مصطلحات البحث، وبذلك فإن استرجاع **الكتب أو الدوريات**، يعني أن الباحث يرغب في استرجاع كل التسجيلات المتعلقة بكل مصطلح من المصطلحات الواردة في صيغة الاستفسار سواء جاءت تلك المصطلحات منفصلة أي مستقلة بذاتها في التسجيلات، أم اجتمعت مع بعضها.

ت. not (ماعدا): ويطلق عليها البعض "ليس" أو "غير" وهي للرفض أو الاستبعاد negation ، حيث يستبعد هذا العامل من النتائج مجموعة التسجيلات الخاصة بمصطلح البحث الذي يسبقه هذا العامل، وبذلك فإن استرجاع **الكتب ما عدا الدوريات**، يعني أن الباحث يرغب في استرجاع جميع التسجيلات المتعلقة بمصطلح الكتب؛ إلا أنه لا يريد التسجيلات التي ورد فيها مصطلح الدوريات مع مصطلح الكتب.

وكما هو ملاحظ، فإن العامل استبعد التسجيلات التي تضمنت مصطلح الدوريات، بما في ذلك تلك التسجيلات التي جاء فيها المصطلح مصاحباً لمصطلح الكتب، وبذلك يتبين أن علاقة (و) تؤدي إلى تضييق حدود البحث واسترجاع تسجيلات أقل من التي يمكن الحصول عليها في الحالات الأخرى،

وعلى عكس ذلك فإن علاقة (أو) تؤدي إلى استرجاع أكبر قدر ممكن من التسجيلات[1].

بالرغم من شيوع المنطق البولياني في نظم الاسترجاع، فقد ظهرت له عيوب كثيرة أهمها:

١- إنه يلزم الباحث أن يفهم قواعده التركيبية الخاصة بالتقاطع والجمع والاستثناء حتى يمكن الحصول على نتائج صحيحة .

٢- إنه يفرض كتابة كلمة البحث بدقة مثلما أدخلت في ذاكرة النظام كالمفرد والجمع والمؤنث والمذكر والمشتقات وعلامات التنقيط وعلامات الترقيم والمسافات .

فمثلا : يمكن للنظام أن يسترجع كلمة مكتبة و لكن لا يسترجع مكتبات إذا تم إدخال الأولى .

كما إنه يمكن أن يسترجع كلمة عبد الله بمسافة بين الكلمتين ولا يسترجع كلمة عبد الله بدون مسافة في حالة إدخال الأولى والعكس صحيح .

٣- إنه لا يمكن البحث باستخدام " و " كلمتين متتاليتين في حقل واحد لأن النظام سيتطلب أن يسترجع الوثائق التي ترد فيها هذه الكلمات متتالية بنفس الترتيب ولا يقوم النظام بالاسترجاع الجزئي .

فمثلاً : إذا كتبنا مكتبات ومعلومات في نفس الحقل، فانه يفترض وجودهما معًا في نفس العنوان ولا يسترجع منفردًا أي عناوين للوثائق تحتوي على مكتبات منفردة أو معلومات .

---

(١) Pirkola , ARI . Studies on Linguistic Problems and Methods in the Text Retrieval .- Tampere : University of Tampere , 1999 .- p30-31.

٤- إنه يتطلب دقة كاملة في تكشيف الكلمات وإدخالها في حقل الكلمات المفتاحية أو الموضوع لما يمثله الاسترجاع في حقل العنوان من قصور كامل .

٥- إن المنطق البولياني يتطلب إعادة صياغة البحث بعدة طرق لتوسيعه أو تضييقه وفقًا لعدد نتائج البحث وهذا في حد ذاته يتطلب مهارة من الباحث ومعرفة مسبقة بقواعد تخزين المعلومات .

نماذج للبحث بالمنطق البولياني:

أحد الباحثين يرغب في الحصول على المقالات التي نشرت عن موضوع النشر الإلكتروني باللغة الإنجليزية للمؤلف لانكستر	
نظام الاسترجاع المقترح	Library literature
نموذج البحث بالمنطق البولياني	النشر الإلكتروني OR النشر الرقمي AND اللغة الإنجليزية AND لانكستر

أحد اختصاصيي المعلومات يرغب في الحصول على قائمة بالكتب تحت النشر التي تتحدث عن موضوع الهندسة الوراثية	
نظام الاسترجاع المقترح	Books in print
نموذج البحث بالمنطق البولياني	الهندسة الوراثية OR الجينات

أحد المستفيدين يرغب في الحصول على أحدث ما نشر حول موضوع الثروة السمكية ومزارع الأسماك دون التطرق إلى أسماك الزينة	
نظام الاسترجاع المقترح	Agriculture
نموذج البحث بالمنطق البولياني	الثروة السمكية AND مزارع الأسماك NOT أسماك الزينة AND ٢٠٠٧ -٢٠٠٨

٢ / ١ / ٢ / ٣ / ٢ استخدام اللغة الطبيعية:

وفي هذه الطريقة يقوم النظام باستخراج الكلمات الدالة من السؤال المطروح باللغة الطبيعية والتعرف إلى جذور الكلمات وفق الخطوات التالية :

١) يتم استبعاد كل كلمة من الكلمات الشائعة كالأدوات " من ، على، أن ... الخ " و الكلمات العامة " دراسة ، منشورات ... الخ " بناءً على قائمة تسمى قائمة الاستبعاد Stop List .

٢) يقوم النظام بالتعرف إلى جذور الكلمات آليًّا وحذف الأحرف الإضافية كعلامات الجمع والتأنيث في آخر الكلمة أو الإضافات والاشتقاقات إذ أن الأسلوب الشائع حتى الآن هو أن يقوم الباحث بنفسه بعملية قطع الكلمات "البتر الضمني" وتسمح بعض القواعد باقتطاع الكلمات الأولى فيما يبدو مفيدًا في حالات مثل الكيمياء التي تكثر فيها إضافة أحرف أولى للمصطلحات .ومثال البتر مثل البحث بكلمة (مصر؟) التي تستدعي نتائج مثل (مصريون، مصري، مصريين.. وغيرها).

أما في أساليب ما بعد المنطق البولياني فيكون اقتطاع الأحرف آليًّا من خلال الرجوع إلى قواميس إضافات و قواميس جذور معروفة و يكون الاقتطاع عادة على عدة مراحل كثيرة حتى لا يختلط الأمر بالكلمات . مثال :

إذا أردنا البحث عن مشتقات Compute فلا يجوز أن يكون القطع إلى Com لأن ذلك سيأتي بمشتقات للكلمة لا تتطابق مع الكلمات الموجودة في السؤال .

٣) يتعرف البرنامج على الكلمات ذات الرسم الإملائي المختلف مثل جغرافيا و جغرافية .

و تستخدم لذلك أساليب كثيرة أهمها :

أ- تصحيح الأخطاء المطبعية الشائعة التي تقع في أربع فئات مثل :

- ✓ حرف زائد.
- ✓ حرف ناقص .
- ✓ حرفان تم استبدال موضعيهما .
- ✓ حرف خطأ.

وبالتالي يصحح الخطأ المحتمل وتقارن الكلمة المصححة في قاموس الكلمات فإذا وجدها يفترض أنها المقصودة و تقترح الشكل الصحيح على المستفيد.

ب- مقارنة الحروف المتشابهة صوتيًا لاحتمال أن يكون هناك استبدال بينهما مثل **P , B** في اللغة الإنجليزية أو ث، س في اللغة العربية .

٢ / ١ / ٢ / ٣ / ٣ استخدام النص الفائق:

ارتبط استخدام مصطلح النص الفائق hypertext بظهور الويب ؛ حيث تمثل صفحة الويب الوحدة المكونة لمواقع ومصادر المعلومات الويب، ويتم الانتقال بين صفحات الموقع الواحد أو المواقع المختلفة عن طريق روابط بين هذه الصفحات تسمى الروابط الفائقة. وقد أطلق مصطلح النص الفائق أو الربط الفائق على هذا الأسلوب من البحث، نظرًا لإمكانية القفز والتحرك بعشوائية بين مستويات المعلومات المختلفة. ويعرف قاموس المكتبات والمعلومات على الخط المباشر ODLIS النص الفائق على أنه " أسلوب لعرض واسترجاع

المعلومات الرقمية، التي ترتبط معًا في ملفاتها وبياناتها بمجموعة من الروابط، وهو يختلف عن أسلوب استرجاع المعلومات التتابعي. وتسمح الرسومات أو الإشارات داخل النص والروابط المصاغة بلغة الترميز HTML بتصفح المعلومات تبعًا لحاجة المستفيدين. وليس بالتتابع التي جاءت عليه." ويرتبط بمصطلح النص الفائق مصطلح آخر هو الوسائط الفائقة Hypermedia الذي يعبر عن التحرك والقفز من صورة إلى صوت... وغيرها من الوسائط الأخرى. ويعتمد النص الفائق في تكوينه على شقين أساسيين هما(١):

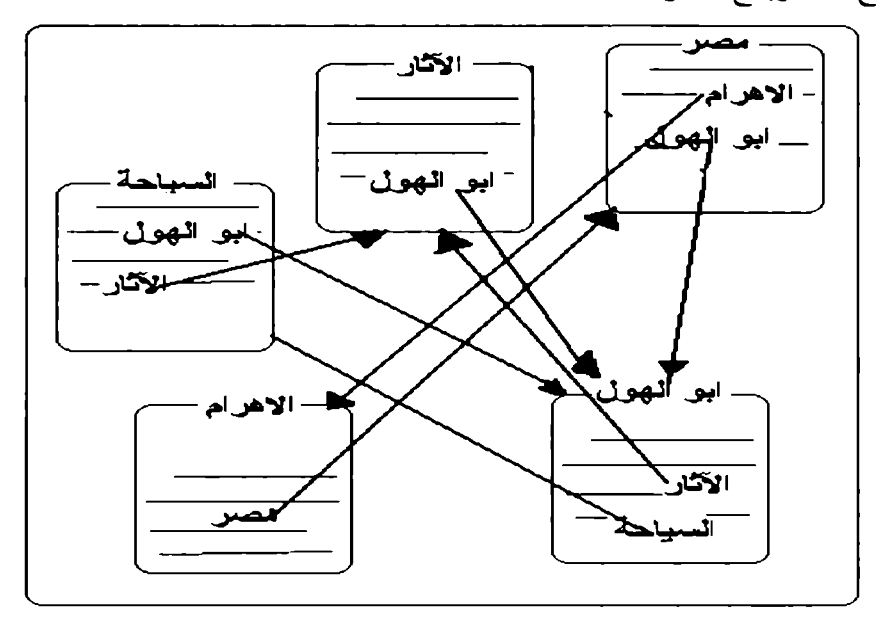
١) **العقد**: nodes وهي عبارة عن الإشارة أو الرمز الذي يعبر عن المعلومات المختلفة التي يمكن للمستفيد الانتقال إليها. وتوجد هذه العقد في صفحات الويب في شكل الكلمات الفائقة أو الوسائط الفائقة.

٢) **الروابط** : Links وهي عبارة عن الوصلات التي تحيل المستفيد من وثيقة إلى أخرى أو من موقع إلى آخر.

يسمح النص الفائق كأسلوب لاسترجاع المعلومات للمستفيد بالتحرك العشوائي بين صفحات الويب، وقد يكون هذا الأسلوب هو أيسر لغات الحوار مع الويب كنظام استرجاع للمعلومات، غير أن أسلوب التصفح قد يأخذ المستفيد إلى التشتت والبعد عن الحاجات الموضوعية المخصصة. وذلك إنما يعود إلى مبدأ التشعب والعشوائية التي يعمل بها النص الفائق، فضلاً عن أن صفحة الويب الواحدة قد تحتوي داخلها على مئات الروابط التي قد تزيد من

(١) Reitz, Joan M. directory, ODLIS — Online Dictionary for Library and Information Science, Libraries Unlimited, 2006, cited 20/8/2007, cited at http://lu.com/odlis/odlis_d.cfm# hypertext

تشتت المستفيد. ويمثل الشكل رقم (٢ -٥) الآلية التي يعمل بها أسلوب التصفح لاسترجاع المعلومات[١]:



الشكل رقم (٢ -٥) تقنية النص الفائق للربط بين صفحات الويب

٢ / ١ / ٢ / ٣ / ٤ أسلوب توسيع السؤال:

تعتمد فكرة هذا النمط من الاسترجاع على أساس أن الباحث غير المتمرس عادة يقوم بتعديل السؤال الخاص به للحصول على نتائج أفضل و هذا يعني أن السؤال الأول الخاص به كان بمثابة تجربة و يحاول الأسلوب المعروف لتوسيع السؤال أن يستفيد من هذه النتائج الأولية للبحث و التقييم الإيجابي من قبل الباحث لوثيقة معينة أو أكثر لاسترجاع وثائق أخرى متشابهة لها فبعد

(١) هانئ محيى الدين عطية مصدر سابق.

أن يطلع الباحث على الإجابات المقترحة لسؤاله الأول و بعد أن يختار الوثائق التي تبدو له ملائمة يقوم البرنامج بتعديل صيغة السؤال الأول و ذلك بإحدى الطرق التالية :

١- أن يضاف إلى السؤال المطروح جميع الكلمات الواردة في عنوان الوثيقة المختارة أو مستخلصها ثم يجمع رؤوس الموضوعات أو الواصفات المستخدمة للتعبير عن محتواه .

٢- أن يضاف إلى الوثيقة رقم التصنيف كمقياس بحث إضافي لاسترجاع جميع الوثائق التي تنتمي إلى نفس الموضوع و كان الشخص يقف أمام رف المكتبة .

٢ / ١ / ٢ / ٣ / ٥ أسلوب أفضل مضاهاة:

تعتمد فكرة أفضل مضاهاة على مطابقة الكلمات الواردة في عنوان الوثيقة أو النص مع الكلمات الواردة في السؤال و إعطاء كل كلمة منها وزنًا بحسب أهميتها و ذلك وفق الخطوات التالية :

١- يقارن النظام بين الكلمات الواردة في السؤال و الكلمات الواردة في النص أو العنوان.

٢- تستبعد الوثائق التي لا تحتوي على أي عنصر من عناصر السؤال.

٣- ترتب الوثائق التي تحتوي على عنصر واحد أو عنصرين أو أكثر من عناصر السؤال حسب وزن الكلمات الدالة فيها و يتم حساب وزن الكلمات بإحدى الطريقتين :

❖ في النظم البسيطة تعطى الكلمة قيمة "١" إذا كانت تحتوي على عنصر ورد في السؤال وعنوان أو نص الوثيقة .

وقيمة " صفر " إذا لم تكن في العنوان أو النص وبالتالي يحسب النظام وزن الوثيقة الإجمالي بالنسبة لطلب معين بعدد الكلمات المتوافرة بها من مجموع الكلمات الدالة المطلوبة في السؤال وترتيب الكلمات الأكثر ملاءمة أو مضاهاة إلى الأقل ملاءمة أو مضاهاة .

❖ في النظم المتطورة يقوم الباحث بإعطاء وزن لكل كلمة يعبر من خلالها عن أهميتها بالنسبة له ويحسب وزن الوثيقة ومدى ملاءمتها للطلب بالنسبة إلى عدد الكلمات المتوافرة وأوزانها المختلفة في السؤال و يحسب الوزن كما يلي :

وزن الكلمة أ في الوثيقة X وزن الكلمة أ في السؤال + وزن الكلمة ب في الوثيقة X وزن الكلمة ب في السؤال .

٤- تحسب أوزان الكلمات الواردة في السؤال آليًا متناسبة عكسيًا مع معدل تواتر الكلمة في مجموعة النصوص في تخصص واحد أي أن الكلمة نادرة الاستخدام في مجموعة النصوص يفترض أنها تعبر بدقة أكبر عن حاجة الباحث من تلك الشائعة في الاستخدام .

٢ / ١ / ٢ / ٣ / ٦ آليات البحث باللغة العربية:

تختلف مسميات المصطلحات داخل عملية البحث بحسب الدور الذي تؤديه هذه المصطلحات، ويعني استخدام مصطلح إستراتيجية البحث أن تتكامل مجموعة من آليات البحث مثل المنطق البوليني والبحث الحر معًا لتشكيل

مجموعة كلمات البحث بزاوية محددة للموضوع، أما آليات البحث فهي مجموعة من آليات التعامل مع كلمات البحث مثل وضعها في حالة الربط البوليني أو وضع التقريب أو في حالة البحث الحر، أما معاملات البحث فهي الإجراءات المختلفة للبحث بشكل محدد للكلمة المفردة أو استخدامها لصنع العلاقات بين كلمات البحث، كحالات العمل في And , Or , Near , Adj , ? ,. وتستخدم محركات البحث على الويب مجموعة مختلفة من الآليات تصلح للاستخدام مع مختلف لغات البحث الممكنة لدى المحرك . وتتسم اللغة العربية بخصائص فرضت على محركات البحث استخدام آليات تعامل مع المفردات العربية تعالج بها مشكلات الاسترجاع الآلي للكلمات العربية . وقد تضمن الفصل الثالث نوعي آليات البحث العربية للكلمة الواحدة و آليات البحث بكلمات متعددة . و يعد التحليل الصرفي للغة العربية باستخدام المعاجم الآلية العامل المحوري لتفعيل آليات بحث للكلمات العربية . وعلى هذا فإن محرك بحث صور باللغة العربية يمكن بحثه بآليات البحث التالية :

أولاً : آليات البحث بالكلمة الواحدة :

- ✓ البحث بالتطابق .
- ✓ البحث غير المطابق " البحث باللواصق " .
- ✓ البحث بجذر الكلمة "حل مشكلة الاشتقاق ".
- ✓ البحث بالتشكيل لكلمات المشترك اللفظي .
- ✓ البحث الحر Wild Card Search

- ✓ البحث بحساسية الحالة .
- ✓ البحث بالترادف.
- ✓ البحث بالمتضادات .
- ✓ البحث بالمعاني .
- ✓ البحث بالترجمة .
- ✓ إمكانية التعرف إلى الأخطاء العربية الشائعة في أثناء البحث .
- ✓ البحث بتشابه النطق في اللغة الإنجليزية.

ثانيًا : آليات البحث بكلمات متعددة :

- ✓ البحث في الحقول .
- ✓ البحث بالتقارب .
- ✓ استخدام المنطق البوليني .
- ✓ طرح سؤال للبحث : قائمة توقف عن الأحرف و الأدوات .
- ✓ استخدام علامات التطابق " ".
- ✓ البحث المعقد أو المركب .
- ✓ البحث بمثال (- /+ " " تقدير الأهمية).
- ❖ استخدام معاملات (- +) .
- ❖ تقدير أهمية المصطلحات .

٢ / ١ / ٣ المستفيد:

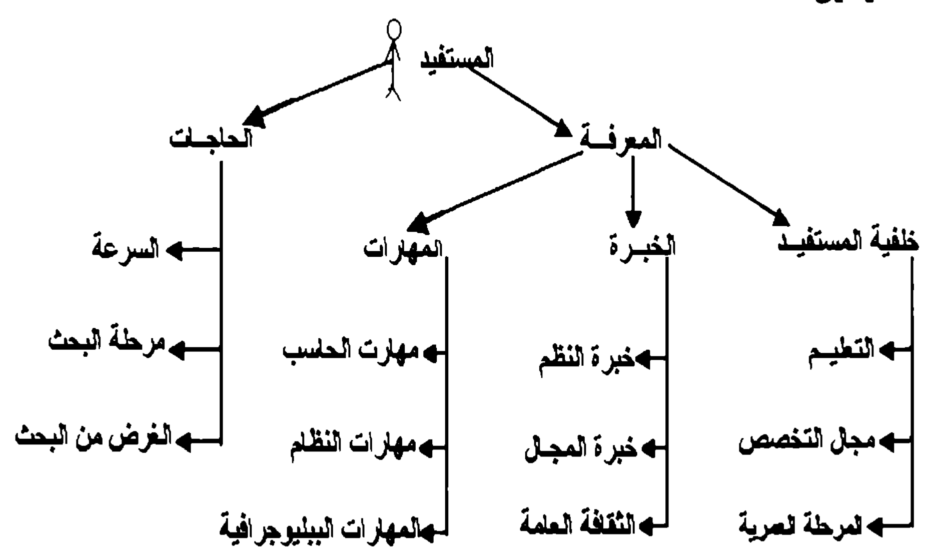
٢ / ١ / ٣ / ١ الملامح المؤثرة في المستفيد:

إن المستفيد هو الغاية التي من أجلها بني نظام استرجاع المعلومات، فعمليات الإضافة للمعلومات والمعالجة الفنية وصياغة لغات البحث إنما تهدف دعم وإمداد المستفيدين بالمعلومات التي تلبي حاجاته. وتكمن أهمية دراسة ملامح المستفيد وسماته في أنه يملك الحكم على نجاح أو فشل نظم الاسترجاع في استدعاء المعلومات المحددة، ولأن نظام الاسترجاع ثابت وموحد في بنيته ولغة بحثه لكل المستفيدين، فما من شك هنا أن عوامل النجاح أو الفشل يمكن أن تترتب على نهج تعامل المستفيدين مع النظام، ومن ثم يمكن القول إن المستفيد ذو القدرات والملامح المحدودة هو الأقل حظًا في التعامل مع نظم استرجاع المعلومات، والعكس من ذلك فإن المستفيدين ذوو القدرات والمهارات العالية هم الأوفر حظًا في الوصول لأعلى درجات التحقيق في استرجاع وثائق نظم استرجاع المعلومات.

ويختلف المستفيدون في التعامل مع نظم استرجاع المعلومات في كل من البيئة التقليدية وبيئة الويب، فالمستفيد في البيئة التقليدية دائما ما يسعى للتعامل مع أدوات النظام كالفهارس وقوائم الاستناد لمحاولة إجادة التخاطب بينه وبين نظام المعلومات، ويساعد المستفيد على ذلك سهولة التعامل مع نظم المعلومات التقليدية، وهذا خلاف ما يكون عليه نظام استرجاع المعلومات الرقمي في بيئة الويب من ديناميكية وتغير سريع ولغات بحث تعتمد على استخدام آليات ومعاملات بحث تحتاج إلى مهارات أعلى وأوفر عند المستفيد.

ويعتمد المستفيد على الوسيط في حالة افتقاده لهذه المهارات سواء في استخدام نظم استرجاع المعلومات التقليدية أو الآلية، ذلك لما يفترض أن يتسم به الوسيط من مهارات تمكنه من الوساطة ونقل الحاجة المعلوماتية من ذهن المستفيد إلى إستراتيجيات بحث نظام المعلومات.

إن طبيعة المستفيد كإنسان ولد ونشأ بملامح محددة (المهارات والقدرات) هي التي تؤثر دائمًا على نوعية التعامل بينه وبين نظام استرجاع المعلومات ؛ حيث يوجد الكثير من العوامل المؤثرة في بنائه فكريًا وفي قدراته على استخدام نظم المعلومات. ويوضح الشكل رقم (٢ -٦) عوامل التأثير في المستفيدين:



العوامل المؤثرة في سمات المستفيدين

الشكل رقم (٢ -٦) يوضح العوامل المؤثرة في سمات المستفيدين من نظم استرجاع المعلومات.

يوضح الشكل رقم (٢-٦) عوامل التأثير في ملامح المستفيدين، وهي كالتالي[1]:

٢ / ١ / ٣ / ١ / ١ أولاً : المعرفة:

يحتوي مفهوم المعرفة هنا على معنى شامل للحالة الفكرية التي يكون عليها المستفيد، والمعرفة هنا تتألف من مجموعة من العناصر التي تبنى بها الخلفية المعرفية للمستفيدين، وهذه العناصر هي:

١) خلفية المستفيد :

يقصد بخلفية المستفيد هنا القدرات العلمية والفكرية التي اكتسبها المستفيد، وهي بدورها تملك قدرًا كبيرًا من الأهمية في نجاح أو فشل التعامل بين المستفيد ونظم استرجاع المعلومات. وتتلخص في :

◆ التعليم :

يقصد بالتعليم هنا الدرجة العلمية التي وصل إليها هذا المستفيد، فكلما علت الدرجة العلمية للمستفيد زادت قدراته الفكرية والعقلية في مخاطبة نظام استرجاع المعلومات. وهو ما يرتبط دائما بالقدرة على فهم معاملات الربط المنطقي في إستراتيجيات البحث، أيضًا يستطيع المستفيد فهم الرسائل الحوارية التي تخرج من النظام، كما يملك المستفيد بدرجته الفكرية العالية ثروة من المصطلحات العامة أو المتخصصة التي تمكنه من تحديد المترادفات والانتقاء بين المصطلحات التي ستكون إستراتيجية البحث.

(١) هاني محيي الدين عطية. مصدر سابق.

◆ مجال التخصص :

تختلف قدرات المستفيدين في التعامل مع نظم استرجاع المعلومات بحسب نوعية التخصص العلمي أو العملي الذي يشغله المستفيد، وهذا يرجع إلى اختلاف التخصصات العلمية والعملية في التفاعل مع النظم والأجهزة بشكل عام ونظم استرجاع المعلومات الآلية بشكل خاص. فمن المعروف أن التخصص في العلوم الطبيعية والتطبيقية يجعل المستفيد في تفاعل دائم مع النظم والأجهزة والمعامل... وغيرها، خلاف ما تكون عليه العلوم الإنسانية والاجتماعية من شبهِ بُعْدٍ تام عن الاحتكاك بالمعامل أو الأجهزة. وعلى ذلك فالأقرب من الأجهزة والحاسبات هو الأقرب من فهم نظم استرجاع المعلومات.

◆ المرحلة العمرية :

إن المقصود هنا هو المرحلة العمرية الصغيرة أو المتأخرة للمستفيدين التي تؤثر في سهولة التفاعل والتعامل مع نظم استرجاع المعلومات. ودائمًا ما يكون المستفيدون الأصغر عمرًا هم الأقدر على التواصل مع نظم استرجاع المعلومات، وهذا يرجع إلى أن الأطفال أو الصغار قد شهدوا ما حدث من تطور تقني في الاتصالات والمعلومات، بخلاف المراحل العمرية الأكبر التي تخشى الاقتراب من الأجهزة والحاسبات نظرًا للارتباط بالطرق التقليدية في الحياة والعمل. ومن ثم يكون الأقرب للتقنيات والأجهزة المختلفة هو الأقدر في التفاعل مع نظم استرجاع المعلومات.

٢) خبرة المستفيد :

تتمثل خبرة المستفيدين في أوجه متعددة ومنها:

◆ **خبرة النظام :**

يختلف المستفيد حديث التعامل مع نظام الاسترجاع عن المستفيد ذي الخبرة في استخدام نفس نظام استرجاع المعلومات. ويعود ذلك إلى إمكانات التمرس والصياغة وآليات البحث التي يستخدمها نظام الاسترجاع، فضلاً عن المهارات المكتسبة من خلال مواجهة المشكلات البحثية السابقة ودراسة رسائل المساعدة التي وجدها المستفيد.

◆ **خبرة المجال :**

تمكن الخبرة في المجال أو التخصص المستفيد من المتابعة والإحاطة الشاملة بكل ما يحيط بمصطلحات ومفاهيم التخصص. وعلى ذلك فإن المستفيدين ذوي الخبرة في المجال هم الأقدر على صياغة إستراتيجيات البحث بدقة واستخدام الكلمات البحثية بما تحمله من مفاهيم محددة، وهذا خلاف الحال الذي يكون عليه المستفيد حديث العهد بالتخصص من ضعف في فهم بعض المصطلحات أو صعوبة في الإحاطة بكل مترادفاتها.

◆ **الثقافة العامة :**

تتسم مختلف نظم استرجاع المعلومات على الويب الآن بالشمولية في تغطية محتوياتها من المعلومات، وهذه الشمولية تحتاج من المستفيدين التعرف إلى المصطلحات العامة والمختلفة إذا ما أراد استخدام نظام الاسترجاع للبحث في مثل هذه الموضوعات العامة. ويمثل هذا العامل أهمية في تحقيق النجاح بين المستفيد والنظام؛ لأن المصطلحات العامة دائمة التغيير، ولا يمكن لأحد المستفيدين إجراء بحث لاسترجاع معلومات عن الزهور دون العلم بالمسمى

الآخر، وعلى ذلك فإن النتائج المسترجعة جميعها ستكون فيما يتعلق بالزهور دون الوثائق المرتبطة بمسمى الورود.

٣) مهارات المستفيد:

يستطيع المستفيد ذو المهارات المتقدمة أن يتفاعل بنجاح مع نظام الاسترجاع أكثر من غيره من المستفيدين الآخرين، وجوانب المهارات تكمن في:

◆ المهارات الببليوجرافية :

لا يمكن لأي من المستفيدين التعامل مع نظم المعلومات التقليدية أو الآلية دون الحد الأدنى من مهارات التعرف إلى بطاقات الفهرسة أو البيانات الببليوجرافية، فكيف يمكن لمستفيد استرجاع التسجيلات الببليوجرافية والبحث في الحقول النشطة داخلها دون معرفة الربط بين هذه الحقول وصياغة إستراتيجية البحث.

❖ مهارات استخدام الحاسب الآلي :

تعمل نظم استرجاع المعلومات من خلال نوافذ الحاسبات الآلية، وبالاعتماد على نظم التشغيل التي يعمل بها الحاسب الآلي. وعلى ذلك فإن فقدان المهارات اللازمة للتعامل مع الحاسبات الآلية يجعل المستفيد غير قادر على الاقتراب من نظم استرجاع المعلومات، ناهيك عن الحديث عن استخدام إستراتيجيات وآليات ومعاملات البحث.

◆ مهارات نظام الاسترجاع :

يملك كل نظام لاسترجاع المعلومات ملامح خاصة سواء أكان ذلك في لغات البحث المستخدمة أو آليات البحث التي تصاغ بها إستراتيجيات البحث.

ومن ثم قد يملك المستفيد مهارات استخدام نظام بعينه نظرًا لكثرة التفاعل معه دون أن يملك المهارات نفسها في التعامل مع النظم الأخرى. والتفاعل الناجح بين المستفيد والنظام دائمًا يترتب على مهارات فهم المستفيد لآليات النظام.

٢ / ١ / ٣ / ١ / ٢ الحاجة للمعلومات:

تؤثر طبيعة حاجات المستفيدين من المعلومات على درجة النجاح في التفاعل مع نظام استرجاع المعلومات، حيث يمكن أن تشكل هذه الحاجات عاملاً ضاغطًا على المستفيد يجعله يفقد القدرة على تحديد المصطلحات أو سوء الاختيار بين النتائج المسترجعة. ويتمثل تأثير الحاجات الموضوعية على المستفيد في العناصر التالية:

أولاً : السرعة :

يرتبط نجاح المستفيد في الحصول على المعلومات المحددة من نظام الاسترجاع على مجموعة من العوامل، ومن أهمها الوقت المتاح أمام المستفيد للاختيار والانتقاء وإجراء عمليات بحث مختلفة للوصول إلى النتائج المرضية. غير أن التسرع في الحصول على المعلومات قد يؤثر على مدى تحديد المستفيد لما يلائمه من آليات وكلمات بحث، وينعكس ذلك عمليًّا على النتائج التي يسترجعها نظام الاسترجاع.

ثانيًا : مرحلة البحث :

تؤثر مرحلة البحث التي يمر بها المستفيد على قدرته في اختيار مصطلحات البحث، فمن المعروف أن الباحث في بداية بحثه يرغب في الحصول على كل ما له صلة قريبة أو بعيدة بموضوع البحث. خلاف ما

يكون عليه الباحث في المراحل النهائية للبحث من احتياجه فقط لمعلومات محددة عن نقاط موضوعية محددة. ويلجأ الباحث في كل من الحالتين إلى اختيار الكلمات البحثية التي تلائم ذلك، وهو ما يشكل بدوره طبيعة النتائج المسترجعة.

ثالثًا : الهدف من البحث :

تتنوع أهداف المستفيدين من الإقبال على نظم استرجاع المعلومات واستخدامها. فمن المستفيدين من يريد الحصول على معلومات بهدف إثراء الثقافة العامة، ومنهم من يريد الوصول إلى المعلومات المتخصصة المحكمة. وبين هذا وذاك فإن النتائج المسترجعة وطبيعة التفاعل بين المستفيد ونظام استرجاع المعلومات تأخذ شكلاً مختلفاً.

٢ / ١ / ٣ / ٢ التفاعل بين النظام والوسيط والمستفيد:

مستويات الحاجة لدى المستفيد تتمثل في أربعة مستويات هي:

- **مستوى الحاجة الكامنة:**

وفي هذه الحالة تكون الحاجة المعلوماتية مبهمة لدى المستفيد، فهناك احتياج ما للمعلومات كامن في اللاشعور غير أن المستفيد لا يستطيع أخذ القرار بالتعرف إلى طبيعة هذه المعلومات المفيدة له.

- **مستوى الحاجة الواعية**

وهي المرحلة التي يستحضر فيها المستفيد ملامح الحاجة المعلوماتية غير أنه لا يزال غير قادر على التعبير عنها.

- **مستوى الحاجة المصورة :**

وهـي المـرحلة الـتي يـتراءى للمـستفيد فـيها ملامـح الحاجـة المعلوماتـية والاحتياج لمعلومات عن موضوع أو مجال محدد ، غير أن القدرة على تحديد المصطلحات الدالة أو التعرف إلى مسار البحث الصحيح لا يزال غائبًا عن المستفيد.

- **مستوى الحاجة القابلة للتداول:**

وهي المرحلة التي يستطيع فيها المستفيد التعرف إلى ملامح احتياجاته من المعلومات، ويتعدى ذلك إلى القدرة على تداول المصطلحات ومناقشة الكلمات الدالة الملائمة لبحث واسترجاع المعلومات مع الوسيط أو أمام النظام.

٢ / ١ / ٣ / ٣ العوامل المؤثرة في تفاعل المستفيد مع الوسيط :

يمكن القول إن هناك مجموعة من العوامل التي تؤثر في نوعية التفاعل بين المستفيد والوسيط في بحث واسترجاع المعلومات داخل نظم الاسترجاع. ومن أهم هذه العوامل:

١- قدرة المستفيد على تحديد الاحتياجات الموضوعية والتعرف إلى المصطلحات الملائمة أو غير الملائمة لبحث واسترجاع المعلومات. وهذه المسألة ترتبط بالخبرة العملية والمعرفة ودرجة التخصص التي يوجد عليها المستفيد.

٢- مهارة الوسيط كهمزة وصل بين المستفيد ونظام الاسترجاع؛ حيث يتحمل الوسيط نقل الفكرة أو الأفكار من ذهن المستفيد إلى نظام

الاسترجاع مصاغة في شكل كلمات أو مصطلحات تعبر عن المعلومات المراد استرجاعها.

٣- درجة التوافق في البنية المعرفية بين المستفيد والوسيط؛ فكلما زاد التوافق اقترب المستفيد والوسيط من مصطلحات البحث الأكثر دقة في التعبير عن الحاجات الموضوعية. والعكس من ذلك صحيح.

٤- تؤثر أيضًا التوقعات المسبقة للمستفيد حول نظام الاسترجاع في مدى الإفادة المرجوة أو درجة التخصص في احتواء المعلومات موضع البحث. وإذا ما توقع المستفيد الاستجابة الضعيفة من نظام الاسترجاع حول موضوعه، فإن استخدام المستفيد للمصطلحات الدالة سيكون أبعد كثيرًا عن الصياغة الدقيقة للتعبير عن حاجته الموضوعية.

٥- تؤثر طرق التفاعل بين المستفيد والنظام في القدرة على التوصل إلى المصطلحات الدالة الملائمة لطبيعة الموضوع. تتنوع طرق التواصل بين المستفيد والوسيط بين الاعتماد على الهاتف أو المراسلات الكتابية أو المقابلة الشخصية. ويمكن القول إن المقابلة الشخصية هي أفضل طرق التواصل بين المستفيد والوسيط في تداول ومناقشة الحاجات الموضوعية.

٦- مدى تعقد الموضوع المتداول؛ حيث تؤثر درجة تخصص الموضوع في إمكانية تداوله بين المستفيد والوسيط، ذلك للاحتياج إلى وقت أطول في تفهم المصطلحات ودلالتها.

٧- تؤثر بيئة التفاعل بين المستفيد والوسيط أيضًا على كفاية التواصل؛ حيث يختلف تداول المصطلحات المناقشة بجانب النظام عنه بعيدًا عن النظام. حيث يمكن للوسيط إجراء مجموعة من عمليات البحث التجريبية داخل نظام الاسترجاع ومتابعة النتائج المسترجعة ومن ثم يمكن التعديل في المصطلحات الدالة على الموضوع. وهو ما لا يمكن إجراؤه إذا ما تمت المقابلة بعيدًا عن نظام الاسترجاع.

٨- تساعد أدوات العمل داخل النظام على زيادة قدر التواصل بين المستفيد والوسيط؛ حيث يمكن من خلال قراءة المستفيد لاستمارات البحث داخل النظام أو التصفح في المكانز المتخصصة الوصول إلى الشكل الأفضل من مصطلحات البحث المعبرة بدقة عن الحاجات الموضوعية.

٢ / ١ / ٤ التكامل بين عوامل بناء نظم الاسترجاع:

ويرى الباحث في ختام هذا الموضوع أن المراحل التي تدور في فلكها العوامل السابقة داخل نظم استرجاع المعلومات الآلية تمر بخطوات كثيرة، تبدأ بمرحلة إضافة الوثائق وانتهاءً باستجابة النظام لمطالب المستفيدين تأخذ التسلسل الآتي[١] :

١- حال ما يتم استلام الوثائق المعدة للإدخال تجرى عملية التحليل الموضوعي لها والتي يقصد بها توزيع محتوى الوثائق إلى مفاهيم موضوعية محددة .

(١) طلال ناظم الزهيري. مصدر سابق.

٢- تجرى عملية ترجمة المفاهيم الموضوعية بالتعبير عنها بمصطلحات موضوعية دقيقة.

٣- تجرى عملية تخزين المعلومات الببليوجرافية لكل وثيقة في قواعد البيانات المعدة لهذا الغرض، وقد يضاف لها مستخلص أو يتم تخزين كامل نص الوثيقة، و يعتمد ذلك على نوع قاعدة البيانات.

٤- يتم إدخال المصطلحات التي تم إعدادها في مرحلة سابقة لكل وثيقة في حقل خاص بها بنفس التسجيلة التي تخزن فيها الوثيقة .

٥- يتم تحويل كل أو جزء من مفردات الوثيقة التي وردت في النص إلى مصطلحات كشفية باستخدام تقنيات التكشيف الآلي ، وفي الوقت نفسه يتم استبعاد جزء من المفردات التي ليس لها معنى أو تأثير في المحتوى الموضوعي للوثيقة.

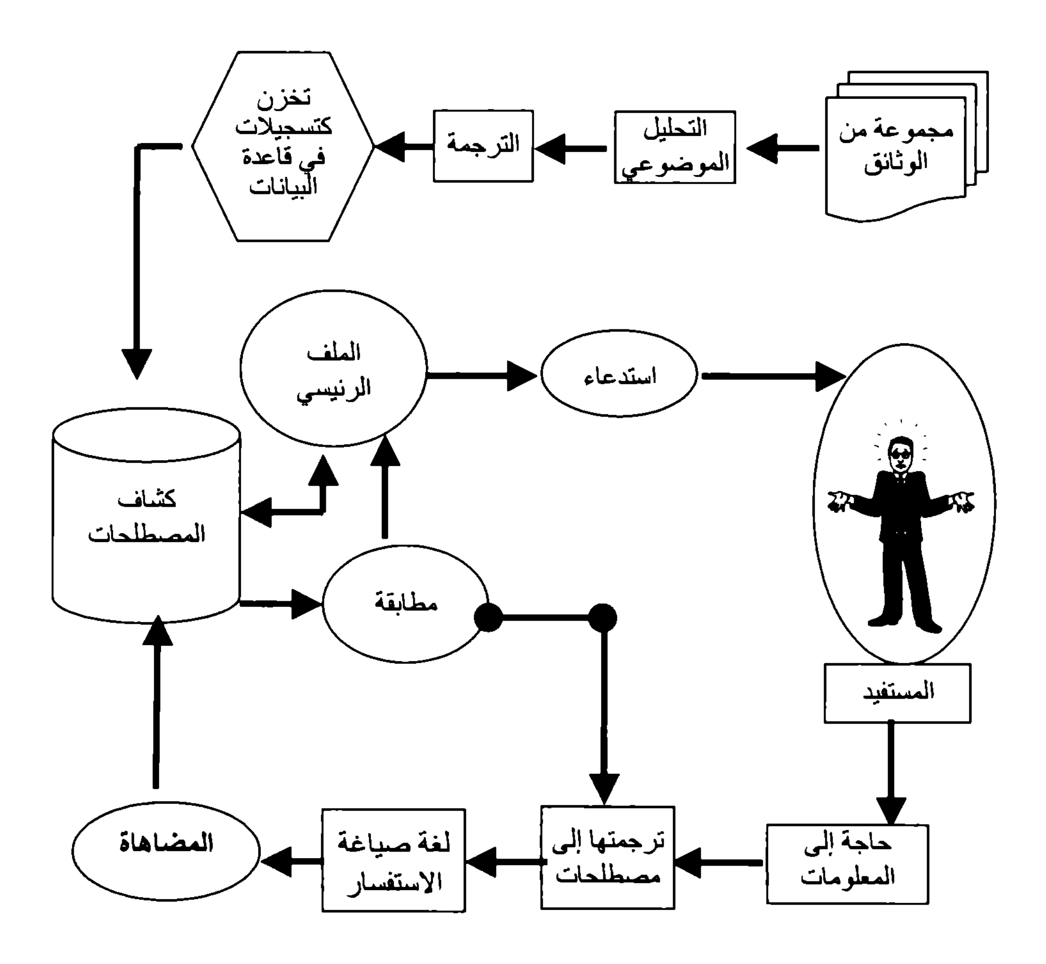
٦- يقوم نظام إدارة قواعد البيانات في بناء كشاف بالمصطلحات الذي يسمى أحيانا بالملف المقلوب (Inverted file) والذي تم اختياره من محتوى الحقول المنتخبة وبحسب آلية أو تقنية التكشيف المختارة .

٧- تجرى عملية الربط بين المصطلح الموجود في كشاف المصطلحات وبين التسجيلات التي ورد فيها .

٨- تتم عملية الاسترجاع عندما يكون المصطلح أو مجموعة المصطلحات التي قدمها المستفيد لها وجود حقيقي في كشاف المصطلحات أو الملف المقلوب

٩- من مجموع محتوى قاعدة البيانات يتم استدعاء التسجيلات التي تحقق فيها الربط مع المصطلحات الموجودة في الكشاف .

ولكل من هذه الخطوات علاقة مباشرة بكفاية الاسترجاع وقدرة تعلق الأمر بنظام استرجاع المعلومات المستخدم في استجواب قواعد البيانات.



الشكل رقم (٢-٧) تكامل عوامل بناء نظم استرجاع المعلومات لأداء مهام النظام

يوضح الشكل رقم (٢-٧) التكامل بين عوامل اللغة والمستفيد والوثيقة في إتمام عمليات استرجاع المعلومات داخل النظام. وقد عبر عنها الشكل كالتالي:

- **الوثيقة** : تم تمثيلها في مجموعة الوثائق التي تنتقل إلى عملية التحليل الموضوعي، والتعبير عنها بمصطلحات ذات دلالة موضوعية ثم اختزانها داخل ملف الكشاف وقاعدة البيانات.
- **المستفيد** : وهو عامل الدعوة إلى استرجاع المعلومات تبعًا لما يحمله من احتياجات موضوعية، يتم صياغتها في شكل مصطلحات دالة، يسترجع نظام المعلومات ما يضاهيها من وثائق داخل النظام.
- **اللغة** : يعبر الشكل عن عامل اللغة في مرحلة الترجمة ومرحلة صياغة الاستفسار ؛ حيث تتمثل لغة التحليل الموضوعي في المكنز أو قائمة رؤوس الموضوعات المستخدمة داخل النظام في أثناء التنظيم، أما لغة الاستفسار فهي تعتمد على هذه الأدوات أيضًا بجانب استخدام أدوات الربط بين الكلمات لصياغة إستراتيجيات البحث.

٢ / ٢ الخاتمة :

تحتاج نظم استرجاع المعلومات إلى مجموعة من الدعائم لبنائها. ولا يمكن لنظام استرجاع المعلومات البناء أو أداء مهامه إذا ما غاب أحدها. وهذه العوامل تمثلت في اللغة التي يتحاور بها النظام مع المستفيد، والوثيقة التي تمثل أساس بناء نظام المعلومات، والمستفيد الذي يمثل هدف نظم استرجاع المعلومات منذ بدايته للوصول إلى تلبية احتياجاته من المعلومات. وبجانب ذلك فقد تضمن الفصل تقسيمًا مقترحًا لأنواع نظم استرجاع المعلومات اعتمادًا على بيئات وخصائص ووثائق نظم استرجاع المعلومات.

الفصل الثالث

التفاعلية والثبات في نظم استرجاع المعلومات

تفاعلية نظم استرجاع المعلومات :

٣ /٠ التمهيد :

تتنوع نظم استرجاع المعلومات فيما بينها من حيث طبيعة وشكل المعلومات المعالجة داخلها، إلى جانب بعض جوانب التنوع الأخرى المتمثلة في أساليب البحث والبيئة التي يعمل بها نظام استرجاع المعلومات. ويهتم هذا الفصل من الدراسة بالتركيز على أحد جوانب الاختلاف بين نظم استرجاع المعلومات، ألا وهو الثبات والديناميكية في عمل نظم استرجاع المعلومات. وعندما نتحدث عن التفاعلية أو الديناميكية هنا فإننا لا نقصد التفاعل مع المستفيد فقط، وإنما أيضًا التفاعل الداخلي المتمثل في التغيير والتحديث الدائم للمعلومات التي يعالجها نظام الاسترجاع. ويؤثر جانب الثبات والتفاعلية على المعلومات على اختلاف نظم الاسترجاع فيما بينها من حيث منهجية العمل وأساليب المعالجة المعتمدة وآليات البحث التي تستخدمها هذه النظم، بل ويتعدى الأمر في ذلك إلى طبيعة الخدمات الأخرى التي يقدمها نظام استرجاع المعلومات لمستفيديه.

وقبل الخوض في ملامح الحركة والديناميكية داخل الويب، لابد من الوقوف أولا على مفهوم التفاعلية Iteraction. فقد أورد قاموس المكتبات والمعلومات على الخط المباشر ODLIS هذا المفهوم أسفل مصطلح Interactive وهو عبارة عن " واجهة للحاسب الآلي تم تصميمها لوضع نموذج إدخال للبيانات يلائم العنصر البشري، وهذا النموذج مكون من أوامر وبيانات. كما أنه عبارة عن واجهة تستخدم للتحاور مع المستفيدين في برامج الحاسب الآلي. أيضًا عبارة عن نظم عالية التفاعل مثل برامج ألعاب الحاسب الآلي، حيث

يمكنها تنفيذ أوامر ورغبات المستفيدين من البرامج." وينطبق المفهوم السابق على حالة واجهات ونظم استرجاع المعلومات على الويب، بل وينطبق على الويب عامة كنظام كوني لاسترجاع المعلومات. بمفهوم التفاعلية يعكس دائمًا التحاور والتخاطب بين رفين أحدهما يتمثل في المستفيد والثاني يتمثل في نظم الاسترجاع[1].

٣ / ١ خصائص الويب التفاعلية والديناميكية :

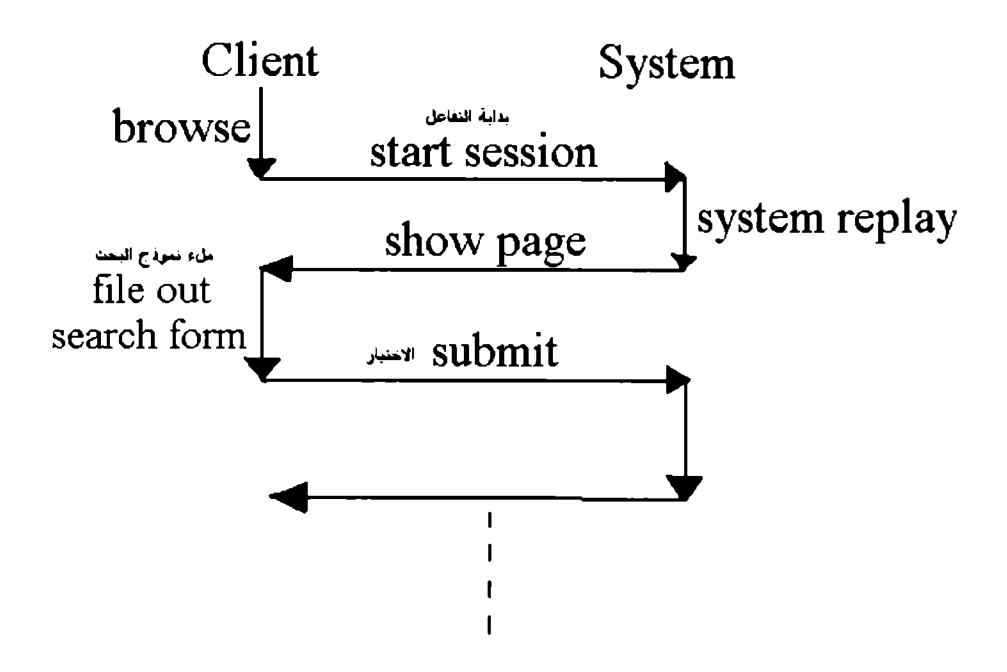
تنطلق مسارات المعلومات الويب في فضاء مترامي الأطراف تتقابل فيه وتتباعد بشكل عشوائي، يعتمد فقط على التقابل عند الارتباطات الموضوعية، ويتباعد عند تشتت وتنوع الموضوعات التي تحملها الويب. وبين مرحلتي بنية المعلومات واسترجاعها تتم عمليات التفاعل المختلفة بين معلومات الويب ونظم استرجاعها، وبين نظم الاسترجاع والمستفيدين، وهذا فضلاً عن الديناميكية والتفاعلات التي تتم داخل نظم استرجاع المعلومات ذاتها. وإن جاز لنا التشبيه هنا، فإن الباحث يرى الويب فضاءً مليئًا بالأجسام صغيرة وكبيرة الحجم، سريعة الدوران، متراكبة ومتقاطعة المسارات، متقاربة ومتباعدة عن بعضها البعض، تتكتل أحيانًا وتتفتت أحيانًا أخرى. هكذا هي الويب ومعلوماتها الرقمية.

تتكون الويب من مجموعة من الخادمات web servers تحمل مجموعات المعلومات الرقمية، ولكي تصل هذه المعلومات من مصممي الويب إلى

---

(١) Reitz, Joan M. ODLIS —Online Dictionary for Library and Information Science, 2006, cited at 1/12/2007, cited at: http://lu.com/odlis/odlis_e.cfm# interactive

مستخدميها، فإن هناك الكثير من التفاعلات بين البرامج وبعضها البعض، وبين برامج نظم الاسترجاع وبين المستفيدين. ولا يمكن تخيل التفاعل بين النظام والمستفيد يتم بطريقة عشوائية كتلك التي تخزن أو تسترجع بها المعلومات، وإنما يتم الحوار بن النظام والمستفيد في تتابع منطقي ومحكم من الأوامر وإجراءات التنفيذ. فعندما يبدأ المستفيد بتصفح الويب والوصول إلى النظام الملائم لموضوع البحث، فهو بذلك قد قصر حواره على نظام الاسترجاع. ويعبر عن ذلك الشكل رقم (٣-١) الممثل لحوار المستفيد مع النظام.



الشكل رقم (٣-١) تتابع إجراءات الحوار بين المستفيد والنظام

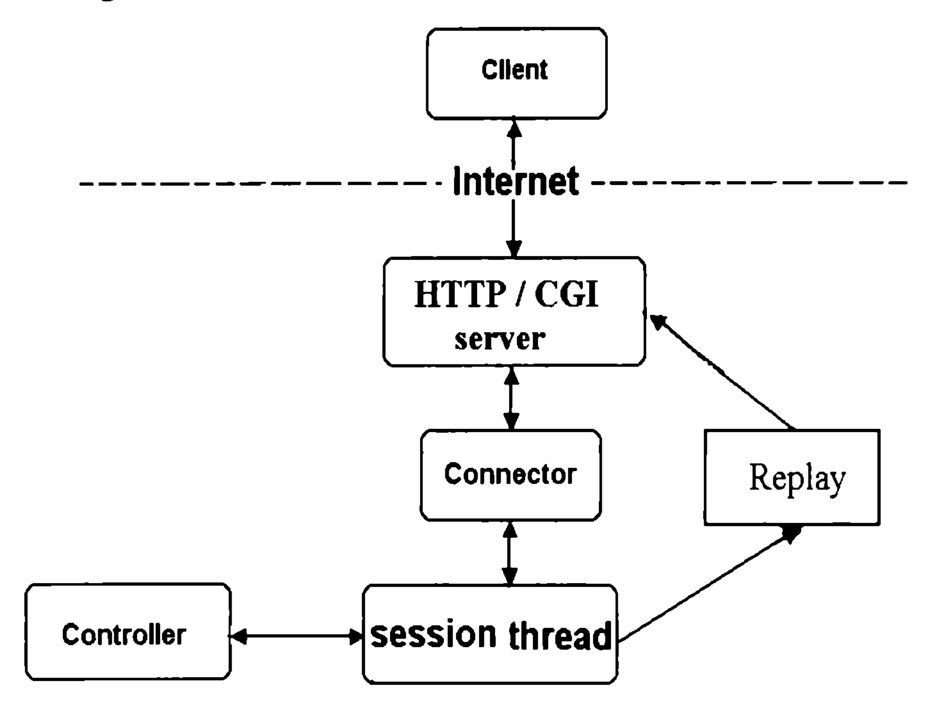
ويسعى الباحث في هذا الجزء إلى إلقاء الضوء على مظاهر وملامح التفاعلية والديناميكية داخل الويب، متمثلاً في العناصر التالية :

٣ / ١ / ١ ديناميكية بنية الويب dynamic web structre:

اعتمدت الويب منذ بدايتها على التشابك والتداخل من خلال وصلات الاتصال وأجهزة الحاسبات الآلية. وارتبطت الويب ببنيتها بلغة كتابة النص الفائق HTML التي تجعل الخضم الهائل من المعلومات عبارة عن مسارات وخطوط متداخلة ومتوازية، تماما كبيت العنكبوت web. وقد مكنت بنية الويب المستفيدين من التفاعل اللحظي مع صفحات الويب؛ حيث يرى المستفيد نتيجة تصفحه وتتبعه لمسار الصفحات في نفس الوقت. وقد عبر Claus Brabrand عن ذلك بتناوله لمخطط التفاعل بين المستفيد ونظم الاسترجاع في طرح استفسارات البحث وتلقي الإجابات في شكل نتائج مسترجعة، ويتمثل هذا التفاعل في الترجمة بين محارف لغة ترميز النص الفائق وبين واجهات CGI لإيجاد التفاهم بين المستفيد وخادم الويب، وقد أطلق على هذا النظام اسم runtime system. متمثلا في الشكل رقم (٣-٢)[1]:

---

(١) Brabrand, Claus and Møller, Anders. A Runtime System for Interactive Web Services, University of Aarhus, Denmark, 2000, cited at 1/12/2007, cited at http://www8.org/w8-papers/3b-web-doc/runtime/runtime.html



الشكل رقم (٣ -٢) التفاعل داخل نظام وقت التشغيل **The runtime system**

يوضح الشكل رقم (٣ -٢) مراحل التفاعل الداخلي بين المستفيد والنظام وبين أجزاء النظام وبعضها البعض، وهو ما يعبر عن الديناميكية الداخلية لنظم استرجاع الويب في أثناء إجراء عمليات بحث واسترجاع معلومات الويب. ويتكون نظام وقت التشغيل من مجموعة من برامج العمل الداخلية هي:

- **عملاء الويب** :Web clients وهم مستخدمو الويب والمستفيدون من خدمات استرجاع معلومات الويب.
- **خادم الويب** : The HTTP/CGI server ويقوم هذا الخادم باستقبال استفسارات البحث من متصفحات الويب، وواجهات CGI وتحويلها إلى

قاعدة أو مستودع البيانات. ثم استرجاع المعلومات المطابقة لموضوع البحث، وصياغتها وفقًا لوجهة متصفح الويب.

- **مسارات المحادثة : Session threads** حيث يسير الحوار بين المستفيد والنظام في شكل إجراءات متتابعة مثل طرح الاستفسار، التحويل إلى قاعدة البيانات، إجراء المضاهاة، تلقي الإجابات، عرضها أمام المستفيد.
- **الموصلات : Connectors** حيث تنشط هذه الموصلات عند طرح المستفيد لاستفسار البحث، ثم ينشط برنامج التحكم لتحديد ما إذا كانت هذه أول عملية بحث من جانب المستفيد أم لا، فإذا ما كانت العمليات مستمرة عمل برنامج التحكم على ربط العمليات ذات الصلة الموضوعية ببعضها البعض.
- **صفحة الإجابات : Reply page** يعمل نظام الاسترجاع على تخصيص ملف يحوي مختلف المعلومات حول عمليات بحث النظام، والإجابات المسترجعة، بما يمكن للمستفيد معه قراءة هذا الملف في أي وقت. مع تكفل النظام بتحديث هذا الملف دوريًا.
- **برنامج التحكم : The controller** وهو جزء محوري في التفاعل بين المستفيد والنظام، ويشرف على تحقيق التواصل وتلقي الاستفسارات، واسترجاع النتائج أمام المستفيد.

٣ / ١ / ١ / ١ ديناميكية محتوى معلومات الويب dynamic web information

تذخر الويب بشتى أنواع المعلومات ومواقع الويب من هيئات بحثية إلى مراكز علمية ومواقع ترفيهية وأدوات بحث... وغيرها من المواقع التي تموج

بمحتوى متقلب متغير بين وقت وآخر. ومن أهم المواقع التي يتغير محتواها وتتغير أساليب عرضه هي مواقع المتاحف ومواقع التسويق ؛ فدائمًا ما تحتاج المتاحف إلى عرض ما يستجد من قطع مضافة إليها، كما تحتاج مواقع التسويق إلى عرض السلع المتباينة والأشكال المختلفة للسلعة الواحدة. ومن ثم كانت الحاجة ملحة للتعامل مع نمط مغاير لما تكون عليه صفحات HTML من ثبات في المحتوى، أي الحاجة إلى تغيير تلقائي يستطيع نظام الاسترجاع فرضه على المعلومات المختزنة لتغيير نمط وحالة مشاهدتها من جانب المستفيدين.

وقد عملت بعض الجامعات ومعامل الأبحاث على تطوير نمط عرض المعلومات ومد نظم استرجاع المعلومات على الويب بالقدرة على تغيير محتوى وشكل معلومات الويب داخل مستودعاتها. ومن بين المشروعات ؛ مشروع Diet Web الذي تتبناه جامعة ميريلاند University of Maryland، ويهدف هذا المشروع إلى تخليق وثائق ديناميكية Dynamic documents من خلال تحويل محارف البرمجة التقليدية للغة CGI إلى صفحات html باستخدام لغة محارف قابلة للتوسع extended tag language، وهو المفهوم الذي تم تطويره لاحقًا تحت اسم لغة الترميز القابلة للتوسع[١].

وتقوم فلسفة هذا المشروع في تخليق وثائق الويب الديناميكية على إجراء تغيير في امتدادات محارف لغة html tags والتي تستخدم واجهات CGI لعرض

(١) Chancogne, David and Austin, Mark. Web Dieting -- Dynamic and Interactive Environment Development Tools for the Web, University of Maryland, 1998, cited at 12/12/2007, cited at
www.isr.umd.eduzSz~doczSzDietzSzdietweb.pdf/web-dieting-dynamic-and.pdf

صفحاتها ووثائقها، ويتم تغيير محارف الـ html بمحارف أخرى قابلة للتوسع extended tags ؛ حيث يمكن للمستفيد من خلالها معالجة المحتوى الداخلي لوثائق الويب والتحكم به. ومن بين محارف html التي يتم تغييرها ما يلي:

١- حقل %variable% The notation الذي يعطي للمستفيد إمكانية الوصول إلى صفحات الويب المصاغة بلغة html، ويعرف تحت اسم <VAR> tag.

٢- مقدمة حقل tag <COND> التي تعطي المستفيد إمكانية اختيار وتحديد الجزء المراد طباعته من صفحة الويب.

٣- حقل tag <LANG> الذي يعمل على توضيح ومعالجة اللغات المختلفة التي كتبت بها صفحة الويب.

٤- حقل tag <LOOP> وهو الحقل الداعم للربط بين أجزاء صفحة الويب.

٥- حقل tag <MAIL> وهو الحقل الذي يسمح بإرسال رسائل البريد الإلكتروني كرابط من داخل صفحات الويب، مما يمكن أن يحقق التغذية المرتدة لنظم الاسترجاع.

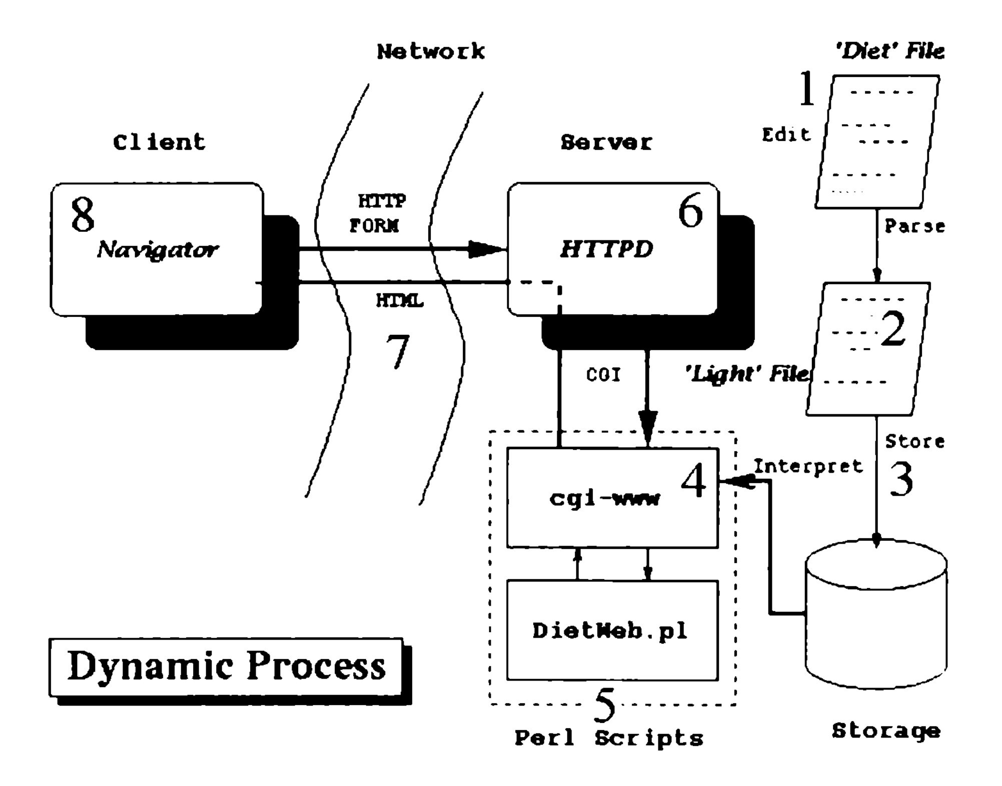
وعلى ذلك يمكن القول إن المراحل التي يتم بها تخليق وثائق الويب الديناميكية هي كالتالي:

أولا التحرير : حيث تصاغ محارف html بالمحارف القابلة للتوسع.

ثانيا التعريب Parsing : وهو ترجمة وصياغة المحارف السابقة لوثيقة الويب إلى لغة perl الداعمة لبناء واجهات العرض العامة CGI.

ثالثًا التفسير interpreting : حيث تنقل وثائق الويب ذات المحارف الموسعة إلى واجهات العرض العامة.

ويمثل الشكل رقم (٣-٣) بنية مشروع تخليق وثائق الويب الديناميكية داخل مستودعات الويب.



الشكل رقم (٣-٣) بنية مشروع تخليق وثائق الويب الديناميكية باستخدام محارف قابلة للتوسع

يوضح الشكل رقم (٣-٣) عناصر وإجراءات بنية مشروع تخليق الوثائق الديناميكية، الممثلة بالأرقام التالية:

(١) يعبر عن عملية التحرير التي تتم لملف صفحة الويب واحتوائه على مجموعة محارف لغة html.

(٢) يتم تعريب وترجمة مجموعة محارف html إلى محارف قابلة للتوسع، ومن ثم يتم تخليق بنية الوثيقة الديناميكية.

(٣) اختزان مجموعات ملفات الوثائق المعربة والمترجمة داخل مستودع المعلومات.

(٤) توجيه مجموعات الوثائق الديناميكية إلى واجهات العرض العامة، حيث يتم تغيير محتواها ونمطها بحسب التحديثات المختلفة داخل مستودع المعلومات، وتبديل نمط عرض الوثائق الديناميكية بين قاعدة البيانات ومتصفح الويب.

(٥) تعتمد واجهات العرض العامة في تغيير محتوى وشكل الوثائق الديناميكية على لغة البرمجة perl و مجموعة من محارف وأشكال النصوص المكتوبة scripts.

(٦) يعمل خادم الويب بوصفه نظامًا لاسترجاع المعلومات على الربط بين احتياجات واستفسارات المستفيدين وبين مجموعات الوثائق الديناميكية وتغيير محتواها وأشكالها بواسطة واجهات العرض العامة.

(٧) يعبر عن البيئة التي تغلف نظام استرجاع المعلومات وهي بيئة الويب، حيث تكون رابط التفاعل بين المستفيد وبين نظام استرجاع الويب.

(٨) يعبر هذا الرمز عن المستفيد متمثلا في متصفح الويب.

يعتبر مشروع diet web محور عمل الويب الآن وتغير محتواها، ذلك لأن عامل الاختلاف والتباين هو المسيطر على مجرى تنظيم واسترجاع المعلومات الرقمية داخلها. فتتعامل الويب مع المستفيدين ذوي أغراض استرجاع المعلومات المختلفة، واختلاف المراحل العمرية التي يعيشونها، إلى جانب تنوع المستويات الثقافية... الخ. أيضًا هناك تشتت تام بين موضوعات المعرفة البشرية التي تحويها الويب في شكل ملفات وصفحات ومصادر معلومات، كما تتنوع أشكال ووسائط حمل المعلومات عن هذه الموضوعات. ومن ثم كان من الجدير بالقائمين على الويب إدراك مبدأي التفاعلية الديناميكية اللذين يؤثران في اختزان واسترجاع المعلومات الرقمية.

٣ / ١ / ١ / ٢ ديناميكية مسارات معلومات الويب Url:

يمثل مسار صفحة الويب أحد أهم أساليب الاسترجاع للمعلومات الرقمية، حيث يرتبط هذا الأسلوب باستخدام متصفحات الويب سواء أكان ذلك من خلال كتابة مسار الصفحات في شريط العنوان، أم من خلال النقر والتصفح والتحرك من صفحة إلى أخرى عن طريق هذه الروابط التي تشكل البنية الشبكية المتشعبة بين صفحات ومواقع الويب.

إن استخدام هذا الأسلوب يتأتى من معرفة المستخدم لعنوان أو مسار موقع المعلومات المطلوبة على الويب، وذلك بالحفظ والوصول المباشر لمسار الويب، غير أن أهم خصائص هذا الأسلوب هو أن يكون المستفيد ذا معرفة بالعنوان أو المسار الصحيح للموقع، مع مراعاة كل العلامات والفواصل وتغير أشكال الحروف small or capital letters، إذ أن خطأ بسيطًا في أي حرف من حروف

المسار قد تؤدي به إلى موقع آخر أو إلى صفحة النتائج غير الصحيحة التي تشير إلى عدم وجود هذا المسار. فعلى المستفيد حفظ مسارات مواقع الإنترنت بشكل صحيح وكتابتها بدقة.

وتؤدي مسارات الويب web URLs دورًا مهمًا في تحقيق الديناميكية التي تحياها معلومات الويب، ومن ثم يطلق عليه حينئذ مصطلح المسار الديناميكي dynamic URLs الذي يعرف على أنه " صفحة الويب الديناميكية التي تحتوي على مجموعة من المعاملات parameters تدفع نظام أو خادم الويب إلى إجراء مجموعة من التغييرات على محتوى وشكل الوثيقة". حيث تعطي المعاملات الإيحاء الدائم للكشاف أو قاعدة البيانات باحتياج وثيقة الويب دائمًا التي تغير المحتوى. ويرتبط مسار الويب بعلاقة وطيدة مع برامج الزاحف ؛ حيث يقوم عمل الزاحف على مسارات الويب التي يعتمد عليها كلية في الوصول لصفحات الويب. ويمكن القول إن المسارات أو الروابط الديناميكية تحدث تأثيرًا مباشرًا على كفاية عمل الزاحف ؛ فلكي يكون المسار ديناميكيًا يجب أن يحتوي على معاملات تسبب التغيير الدائم لمحتوى الصفحة، وهذه المعاملات تغير من شكل مسار الويب وتعريفه لدى الزاحف ومن ثم فيخطئ الزاحف دائمًا عند الرجوع إليها مرة أخرى في التحديث[1].

ويوضح المثال التالي نموذجين لمسار ذي طابع ديناميكي وآخر ثابت:

---

(١) Dynamic URLs and Crawlers. Searchenginewatch, 2007, cited 20/12/2007, cited at http://blog.searchenginewatch.com/blog/070815-092740.

المسار الديناميكي

http://www.plumbersurplus.com/category.aspx?Cat=991

المسار الثابت:

http://www.plumbersurplus.com/Cat/Bathroom-Faucets/991.

٣ / ١ / ١ / ٣ ديناميكية أدوات بحث الويب وتداخل محتواها:

تمثل العمليات والأنشطة التي تتم داخل أدوات بحث الويب نموذجًا للتفاعلية والديناميكية في أداء هذه العناصر مع بعضها البعض؛ ذلك لأنها تتعامل مع محتوى سريع التغير والتقلب. وبإمكاننا تعميق النظر في ذلك إذا ما تطرقنا إلى العمليات والأنشطة التي تتم داخل نظم استرجاع الويب(١).

٣ / ١ / ١ / ٣ / ١ الإضافة والتزويد:

تختلف نظم الأدلة البحثية عن محركات البحث في أنشطة الإضافة والتزويد، فالأولى تعتمد على القوى البشرية، والثانية تعتمد على البرامج الآلية. فالأدلة البحثية تستخدم الأفراد في مرحلة الإضافة والتزويد عن طريق تصفح الإنترنت ثم تجميع المواقع التي وقع عليها الاختيار، وفهرستها وتكشيفها لتوضع ضمن قاعدة بيانات الدليل، أما نظم محركات البحث فتعتمد على برنامج الزاحف أو العنكبوت. وهذا يعني أن حجم التغطية في الأدلة البحثية ليس بالقدر الكافي لمواقع الويب المتنامية العدد. فالأدلة تعمل

---

(١) سيد ربيع سيد. **محركات بحث الصور الثابتة على الإنترنت**: دراسة تحليلية.- الرياض: مكتبة الملك فهد الوطنية، ٢٠٠٧. ص ٥٢.

من منظور الأعداد القليلة التي يمكن التحكم فيها وإعدادها بما يحقق للباحث أكبر قدر من الدقة والتحقيق، بينما تعمل محركات البحث على تغطية واسعة للمواقع على الويب.

ومن ناحية أخرى، فإن التزويد والإضافة للأدلة البحثية يكون عن طريق الطلبات التي تقدم من أصحاب المواقع ذاتهم سواء كانت تجارية أو حكومية أو علمية أو شخصية أو غير ذلك من الأنواع الأخرى. أما الإضافة والتزويد في محركات البحث فيكون عن طريق الزاحف الذي يتحرك بين روابط الصفحات المختلفة ومن ثم تحليل كل الروابط داخل صفحات الموقع المجدد ثم زيارة هذه الروابط، إلا أن برامج الزاحف تختلف فيما بينها من حيث مستويات تحليل صفحات المواقع؛ حيث تكتفي بعض البرامج بالمستويين الأول والثاني لصفحات موقع الويب، وتتعدى بعض برامج الزاحف ذلك إلى المستويات الثالث والرابع. وهو ما يؤثر بشكل أساس على أعداد صفحات الإضافة اليومية لبرنامج الزاحف. وهذا يترتب عليه مدة التحديث التي قد تتأخر بعض الشيء. فالتزويد في محركات البحث يتم في مدد متقاربة كل ساعة أو كل يوم، وقد يكون في مدد متباعدة أسبوعيًا أو شهريًا. إلا أنه يمكن اعتبار هذه المدد دورة يتم فيها الزاحف العمل إلى العودة مرة أخرى لمثل هذه المواقع لتبيان ما بها من تحديث، أو تغيرات في متن هذه الصفحات.

٣ / ١ / ١ / ٣ / ٢ الإعداد الفني والتنظيم :

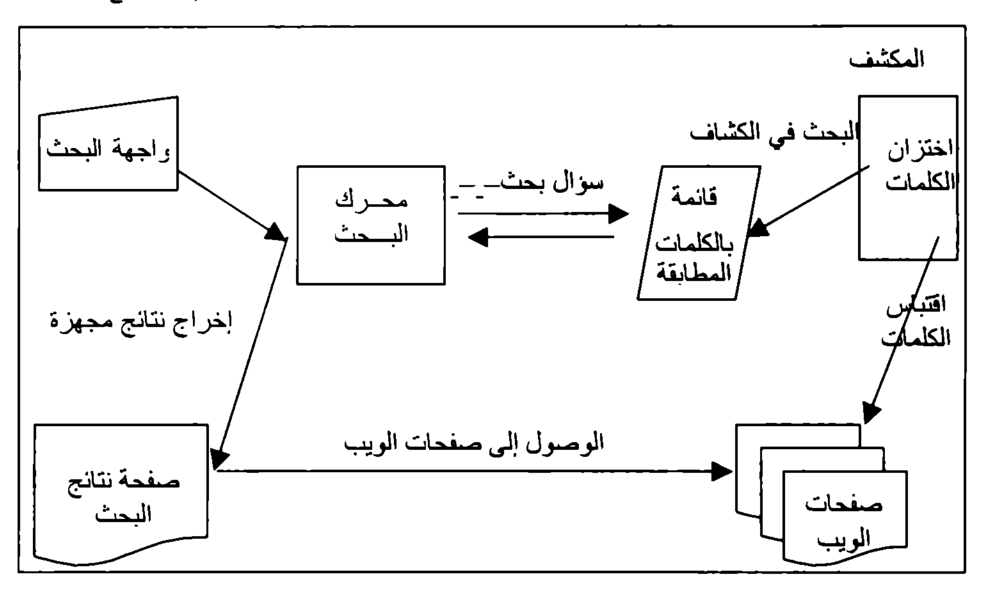
بعد أن تنتهي المرحلة الأولى بإضافة نسخة من موقع الويب إلى مستودع محرك البحث، يبدأ عمل الجزء أو البرنامج الثاني في نظم استرجاع الويب

الذي يعرف بالمكشف أو المفهرس ؛ حيث يقوم بإعداد تسجيلات ببليوجرافية لملفات الويب حديثة الإضافة إلى قاعدة بيانات أداة البحث ، وتعمل هذه التسجيلة وفقًا لما ورد في المكتبات والمعلومات عن (التمثيل المكثف) حيث لا يتم اختزان موقع الويب بأكمله في عدد من أدوات البحث ، وإنما حقول وبيانات منها مسار هذا الموقع للربط به والوصول إليه ، وكل ذلك في شكل تسجيلة كما هو الحال في الفهارس الآلية للمكتبات OPAC. وتتم عملية الفهرسة في حالة الأدلة البحثية بشكل يدوي ، أما في حالة محركات البحث فإنها تتم آليًا ، وفي حالة التكشيف والفهرسة الآلية ، فإن برامج التكشيف الآلي تستخدم بنية صفحة الويب في التعرف إليها ووضع كلمات أو مصطلحات معبرة عن الموضوع أو الموضوعات التي تضمها صفحة الويب؛ حيث يفحص البرنامج هنا كل أجزاء الصفحة التي تحتوي على كلمات تخص الموضوع من كلمات العنوان ، وكلمات النص ، والكلمات المفتاحية التي وضعها مصمم صفحة الويب في رموز meta tags وتجمع هذه الرموز كل الكلمات التي يرى مصمم الصفحة أنها تعبر عن موضوع أو موضوعات صفحة الويب. وبعد تفحص كل الجوانب السابقة يتم وضع كلمات مفتاحية داخل تسجيلة هذا الموقع أو وضع كل كلمات النص الطبيعية في وضع نشط؛ ليتم مضاهاتها بعد ذلك مع كلمات البحث التي يضعها الباحث عند إجراء البحث. وتختلف مواضع التكشيف من برنامج آلي آخر؛ حيث تعمل بعض برامج التكشيف على الاكتفاء بالجزء الأول فقط من بنية اللغة المعيارية لصفحة الويب ، كما تعمل بعض البرامج على تكشيف عناوين الصفحات بجانب

مسار الصفحة، وتمتد قدرة برامج التكشيف في أحيان أخرى إلى تكشيف كامل متن صفحة الويب وهو ما يجعل محركات البحث في هذه الحالة تمتاز بوفرة في الاستدعاء مع انخفاض نسبة التحقيق. ويبين الشكل رقم (٣ -٤) التفاعل والديناميكية التي تتسم بها البرامج الفرعية لنظم بحث واسترجاع معلومات الويب.

٣ / ١ / ١ / ٣ / ٣ البحث والاسترجاع:

يعد الاستدعاء من أهم مقاييس كفاية النظام. لكن الأمر لا يقتصر على النظر إلى النتائج، بل يتعدى إلى ماهية التنظيم داخل نظام الاسترجاع؛ فمحركات البحث تعتمد على البحث الحر في كامل نص الموقع مع إمكانية البحث في العنوان والرموز المعيارية (metatags) وبذلك فإن الاستدعاء سوف يكون في أعلى درجاته، بينما يكون التحقيق في أقل حالاته، ذلك لأن آلية التكشيف التي تعتمد عليها محركات البحث هي استخدام اللغة الطبيعية وهذا إنما يعكس بدرجة عالية أن كل كلمات النص سوف تصبح كلمات مفتاحية يمكن من خلالها استدعاء الوثيقة (موقع الويب) اعتمادًا على مضاهاة الكلمات البحثية. أما الأدلة البحثية فإنها تعتمد في البحث على كلمات كشفية ومصطلحات يتم وضعها بالتدخل البشري تعبر بصدق عن الموضوع. وبالتالي فإنه يستبعد أية وثائق أو مواقع غير ذات صلة، ويولد تحقيقًا عاليًا مع استدعاء منخفض.



الشكل رقم (٣-٤) يوضح تفاعل وديناميكية الأداء للعناصر الفرعية داخل أدوات بحث الويب

ومن ناحية أخرى تكون النتائج في محركات البحث مرتبة حسب أهمية الموقع للموضوع. ويختلف الوزن النسبي لموقع الويب وتقدير أهميته للموضوع إما اعتماداً على عدد تكرار الكلمات المفتاحية التي ذكرت في السؤال، أو أماكن تواجد الكلمات المفتاحية والموقع بحيث إذا ما وردت في العنوان يكون للموقع أهمية أكثر من وجودها في النص نفسه. وبحسب الأهمية يأتي ترتيب النتائج بالموقع فالأكثر أهمية للموضوع يأخذ الرقم (١) في الترتيب ثم ما يليه من المواقع على حسب هذا الوزن ... وهكذا. أما بالنسبة للأدلة البحثية، فإنه عادة ما تحدد هذه الأهمية من قبل المفهرس نفسه الذي يحدد موضع الموقع في الترتيب.

وفي كلتا الحالتين، فإن البحث والاسترجاع يعتمدان على إستراتيجيات تستخدم معاملات المنطق البوليني، وآلية بحث اللغة الطبيعية، والبحث بالتماثل، والبحث بالمفهوم، والبحث بالتقارب بين الكلمات near , next ، والبحث بالمتضادات، والبحث بالترجمة... وغيرها الكثير من معاملات وآليات بحث تعمل جميعا لأداء أفضل تحقيق للباحث عن المعلومات وسيأتي ذكرها بالتفصيل.

إن الأهمية التي تأخذها هذه الأدوات تأتي من الدور الذي تؤديه في الويب باعتبارها جسرًا يهدف الوصول إلى المعلومات المحددة؛ حيث تعمل هذه الأدوات عمل اختصاصي المعلومات في اقتناء وسائط المعلومات في مختلف المجالات ثم تنظيمها وإعدادها الإعداد الفني لوضعها قيد الاسترجاع في أي وقت لمختلف الباحثين وكل حسب حاجته الموضوعية. ولولا هذه الأدوات لأصبحت الويب مستودعًا يحمل بداخله معلومات العالم أجمع في شكل غير منظم وعشوائي لا يوفر أي وسيلة للبحث سوى مسار أو عنوان الموقع URL ، مما ينعكس على المستفيد كالذي يبحث عن ضالة في فلاة[١].

وبذلك يمكن تحديد أهمية وجود نظم بحث واسترجاع معلومات الويب في سببين هما:

١- أن كم المعلومات على الإنترنت ينمو بالسرعة والحجم الذي لا يمكن لمستخدم ما معرفة ما المواقع التي تهمه وأين توجد؟

٢- أن الأغلبية من المستخدمين للويب حتى الماهرين منهم لا يستطيعون الوصول في عمليات البحث إلى مستوى الدقة المناسب والتحكم في عمليات الاستدعاء، ذلك لأنه يبحث عن شيء غير محدد أمامه.

ولا يمكن القول إن محتوى واحدًا من نظم أو أدوات بحث الويب هو محتوى قاصر عليه فقط، وإنما هناك تداخل بين محتوى كل نظام وآخر، ففضاء الويب مليء بملفات المعلومات والصفحات وهو ما يجعلها جميعًا أو معظمها تتداخل ضمن نطاقات أدوات البحث المختلفة. ويأتي ذلك من استقلالية كل نظام عن الآخر ؛ حيث قد يضم مستودع المعلومات في جوجل

(١) المصدر السابق نفسه.

google مثلاً ما يحويه محرك البحث msn، غير أن الاختلاف يأتي فقط في حجم قاعدة البيانات ومستوى التعمق في تحليل الصفحات واختلاف فترات التحديث بين المحركين. مما يعني أن بحثًا عن موضوع ما في google قد يسترجع نفس النتائج في msn مع اختلاف آليات البحث وترتيب وبناء القائمة النهائية للنتائج. ومن ثم فإن ديناميكية التفاعل بين أدوات البحث تأتي في مبدأ تحقيق التكامل بين مستودعات معلوماتها، التي تصل بالمستفيد في النهاية إلى استرجاع أكبر قدر من المعلومات الخاصة بموضوع البحث من فضاء الويب.

٣ / ١ / ٢ التفاعلية والديناميكية في أداء نظم استرجاع معلومات الويب:

تقوم نظم استرجاع المعلومات سواء أكانت التقليدية أم على الويب على مجموعة من البرامج والإجراءات التي تتكامل معًا لتحقيق غاية استرجاع وبث المعلومات للمستفيدين بحسب الحاجات الموضوعية. ومن ثم فإن نظم الاسترجاع على اختلاف أنواعها وتنوع بيئات عملها، تموج بحركة ديناميكية نتيجة تفاعل مكونات أو أجزاء النظام معًا. ومن أبرز النماذج التي تعكس هذه الحركة الديناميكية داخل نظم الاسترجاع، ما ذكره Jun Yang في حديثه عن نظم استرجاع أفلام الفلاش Flash movies، وقد انعكس ذلك على بنية نظام الاسترجاع FLAME (FLash Access and Management Environment) الذي تكون من ثلاث قطاعات أساسية تتدرج من أسفل النظام في مرحلة اقتناء الوثائق، إلى مرحلة استرجاع المعلومات عند المستفيدين، مقسمة إلى أجزاء أخرى متفرعة منها، تمثلت في[1]:.

(١) Yang, Jun , Li, Qing and Wenyin, Liu. Retrieval of Flash™ Movies by Semantic Content: Research Issues, Generic Framework, and Future Directions, City University of Hong Kong, 2001, cited 12/12/2007, cited at www.cs.cmu.edu/~juny/Prof/papers/ieeemm-si-aesthetics.pdf.

أولاً : قطاع الحفظ والتمثيل : representation layer:

حيث يخصص النظام مجموعة من محارف لغات الترميز XML Tags لكي تكتب بها ملفات أفلام الفلاش داخل نظام الاسترجاع Macromedia Flash (SWF) file. وشأن هذه الأفلام في عمليات الإضافة والاقتناء، شأن مختلف ملفات الويب الأخرى. توضع هذه الأفلام مع مختلف أنواع الملفات الأخرى التي يقتنيها النظام، فإذا ما ورد الملف في شكله الثنائي binary لقاعدة البيانات، عمل النظام على إدخاله في شكل ملف نصي باستخدام XML ASCII. ويمكن تحويل بعد ذلك إلى ملف في شكل XML format باستخدام JavaSWF. وتبرز ميزات تحويل مختلف ملفات النظام إلى شكل XML format في.

أ- أن ملفات XML قابلة القراءة بيسر بالشكل الذي يمكننا من فهم بنية المحتوى الداخلي لأفلام الفلاش.

ب- أن استخدام شكل ملفات XML كمعيار عالمي، يجعله قابل المعالجة والاستخدام من قبل البرامج الأخرى.

يحتوي ملف الفلاش على مجموعتين من محارف أو حقول ترميز الملفات وهما ؛ مجموعة محارف التعريف بمكونات الملف defintion tags كحقول التعريف بالشكل والتعريف بالنص، ومجموعة محارف التحكم control tag، التي تقوم بإدارة وتشغيل مكونات الفيلم حتى تعطي إحساسًا كاملاً بالتفاعلية مع المستفيد. وتسمى مختلف حقول التعريف المختزنة داخل النظام بالدليل.

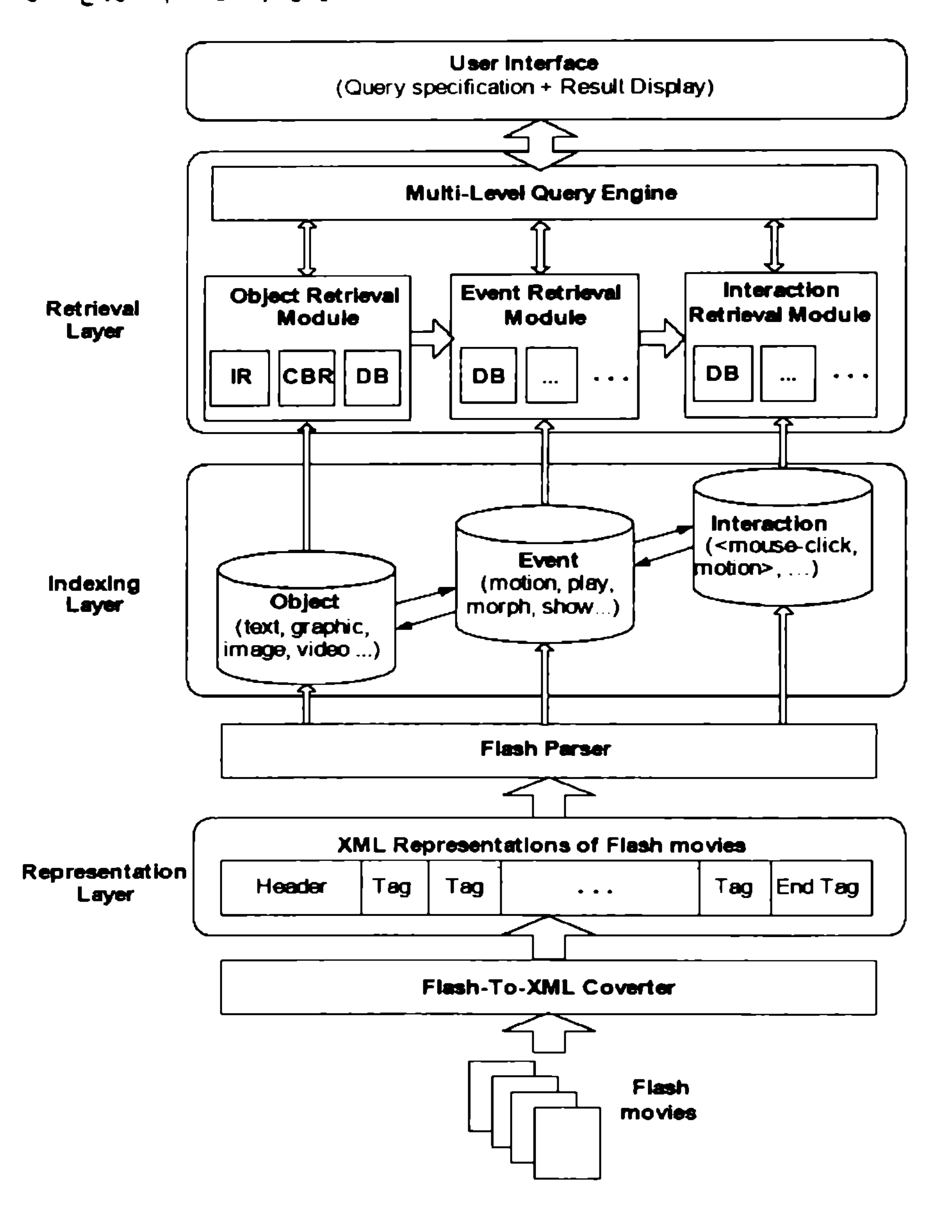


Figure 1: The 3-tier architecture of FLAME

الشكل رقم (٣-٥) بنية التفاعلية نظام استرجاع المعلومات متعددة العناصر التفاعلية

ثانيًا : قطاع التكشيف والتنظيم : indexing layer

يخصص هذا القطاع مجموعة من قواعد البيانات تعمل على تحليل المكونات المختلفة لأفلام الفلاش؛ حيث تقوم إحداها بحفظ وثائق نظام الاسترجاع من أفلام ونصوص وصور، وأخرى تختص بإجراءات التفاعل مع هذه الوثائق وتشغيلها، وأخيرة بالتفاعل مع الكلمات الدالة على محتوى وثائق النظام. ومن يحفظ النظام هذه الملفات على اختلاف خصائصها التفاعلية والديناميكية والبنيوية. ويوضح الجدولان رقما (٣-١) و(٣-٢) خصائص التعامل مع الملفات المتواجدة داخل قاعدتي البيانات **Event , object**.

الجدول رقم (٣-١) الخصائص النوعية للملفات المختلفة داخل النظام

Object	Features
Text	keywords, font size الكلمات الدالة وحجم الخط
Graphic	shape, color, number of borders الشكل واللون وعدد الحدود
Image	size, color, texture الحجم واللون وبنية الصورة
Sound	MFCCs (mel-frequency cepstral coefficients) التجسيم ونوعية مسار الصوت
video	features of a set of key-frames, motion vectors

الجدول رقم (٣ -٢) الملامح التطبيقية للتفاعل مع ملفات المعلومات

Action	Applicable objects	Features
show	all but sound	position, start/end frame
motion	all but sound	trail, start/end frame
rotate	all but sound	angle of rotation, location, start/end frame
resize	all but sound	start/end size, location, start/end frame
morph	graphic	start/end shape, number of frame
play	sound, video	current frame
trace	all but sound	closeness to mouse
navigate	N/A	target URL

ثالثًا: قطاع استرجاع المعلومات: retrieval layer

يختص هذا القطاع بإجراء عمليات البحث وتلقي الاستفسارات بأكثر من شكل، فقد تكون استفسارات البحث معتمدة على الكلمات البحثية وإستراتيجيات البحث، كما تعتمد على الاسترجاع المبني على المحتوى CBCR. ومن ثم تمكن المستفيدين من توجيه الاستفسارات بما يلائم طبيعة المستفيد على التفاعل مع النظام. وكما يوضح الشكل رقم (٣-٥) فإن نظام الاسترجاع يخصص ثلاثة أنماط من المضاهاة هي مضاهاة الموضوعات والأحداث والتفاعل مع ملفات المعلومات، تعمل جميعًا على تحقيق إمكانات الاسترجاع. فعلى سبيل المثال يمكن للمستفيد إجراء عملية بحث عن الأفلام القابلة للتدوير **rotate** ذات الشكل المحدد من داخل مستودع النظام. ومن ثم يمكن للنظام الجمع بين اثنين من خصائص وصف الأفلام.

ويضاف إلى عمل هذا القطاع جانب آخر هو تلقي النتائج المسترجعة وإجراء الترتيب فيما بينها، ثم عرضها أمام المستفيد.

٣ / ٢ ديناميكية نظم استرجاع المعلومات على الويب:

تختلف نظم استرجاع المعلومات على الويب عن غيرها من نظم الاسترجاع الأخرى، ويأتي ذلك من أن نظم استرجاع معلومات الويب تعتمد على أكثر من جانب في صياغة التفاعل بينها وبين المستفيد ومنها:

- ❖ استخدام نظم استرجاع الويب لعوامل التوافق بين منفذ المستفيد وبين خادمات الويب، حتى يسترجع نظام المعلومات نتائجه بما يتوافق وسهولة الرؤية والاستخدام من جانب المستفيد.

❖ سرعة تدوير معلومات الويب ؛ بما يتمثل في تحديث بعضها وإضافة جديد وكسر أو موت بعض الروابط لبعض المعلومات.

❖ التنوع والتباين في أشكال المعلومات التي قد تحملها صفحة أو موقع الويب الواحد ، من معلومات نصية ومصورة ومتحركة... وغيرها.

❖ اتجاه بعض نظم استرجاع المعلومات للتعامل مع المعلومات غير المرئية من الويب المتمثلة في قواعد بيانات الويب التي تعتمد على بنية خاصة وواجهة ذات بنية أخرى.

تقسم نظم استرجاع الويب من حيث الاهتمام بالتحديث والتطوير لمحتوى صفحاتها إلى قسمين هما :

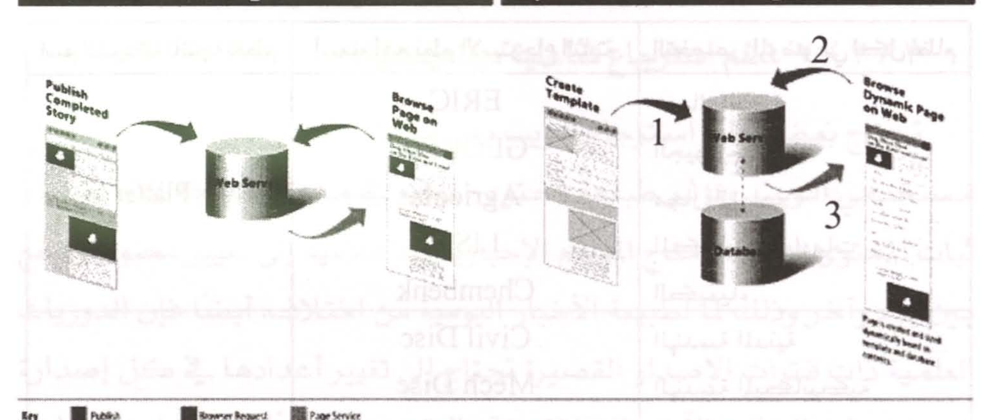
٣ / ٢ / ١ نظم استرجاع ثابتة:

تتصف نظم استرجاع الويب بالثبات أو السكون إذا ظل محتوى صفحات الموقع دون تحديث أو تغيير. أما التفاعلية فهي تغيير محتوى صفحات الويب بما يتوافق مع مستجدات التخصص الموضوعي. غير أن صفة التفاعلية يلحقها بعض مستخدمي الويب بالمصادر التي توفر آليات المساعدة وإتاحة عناصر تشغيل الموقع أمام مستخدمي الويب بما يكفل إمكانية استرجاع معلومات الموقع بيسر تام.

ويوضح الشكل رقم (١-٦) نوعي نظم استرجاع الويب ؛ التفاعلية والثابتة اعتمادًا على التطوير والتحديث في محتوى صفحات المواقع[١]:

---

(١) Timeless Principles of Design (Web Techniques, CMP Technology, 2007, http://www.ddj.com/dept/architect/184414356?pgno=1



الشكل رقم (٣ -٦) نظم استرجاع الويب التفاعلية والثابتة

يوضح الجانب الأيمن من الشكل رقم (٣-٦) صفة التفاعلية في تغيير وتحديث مصدر الويب ؛ حيث يدل السهم رقم (١) على نموذج النشر القابل للتغيير. أما السهم (٢) فيدل على خادم الويب الذي يستقبل آخر التغييرات في محتوى موقع الويب، ويدل السهم ٣ على العمل المتواصل في قاعدة البيانات بإضافة وعرض أحدث تطوير للمحتوى أمام متصفحي الموقع. والجانب الأيسر من الشكل يدل على إدخال معلومات ذات شكل ثابت إلى خادم الويب ثم عرض محتوى الموقع دون الوصول إلى تغيير أو تطويره مرة أخرى.

٣ / ٢ / ١ / ١ نماذج من موردي نظم الأقراص المضغوطة:

لقد أثبتت نظم استرجاع المعلومات الثابتة أو في شكل الأقراص الضوئية قدرتها على تقديم المعلومات للمستفيدين داخل مرافق المعلومات. وقد عملت مختلف شركات إنتاج هذه النظم على متابعة تقديمها وتحديث معلوماتها بحسب الإصدارات المختلفة منها، فضلا عن إتاحة هذه المعلومات للوصول المباشر من خلال الويب. ومن أهم نماذج موردي قواعد البيانات ما يلي:

الجدول رقم (٣-٣) يوضح نماذج لأهم موردي نظم استرجاع المعلومات على الأقراص الضوئية

اسم الشركة المنتجة للنظم	أسماء أهم نظم الاسترجاع الثابتة	التخصص الموضوعي لكل نظام
شركة Silver Platter	ERIC GEOREF Agricola LISA Chembenk	مجال التربية الجيولوجيا مجال الزراعة المكتبات والمعلومات الكيمياء
شركة Dialog	Civil Disc Mech Disc Nuclear science Medline The Philosopher Index	الهندسة المدنية الهندسة الميكانيكية العلوم النووية الطب الفلسفة
شركة Wilson	Applied science ART Index Humanities Index Index to legal periodical Library literrature	العلوم التطبيقية الفنون الإنسانيات كشاف الدوريات القانونية مجال المكتبات
شركة Bowker	Books in Print Ingeram books in print Books out of print Childern books in print Ulrich Plus	الكتب تحت النشر مختلف الأوعية تحت النشر الكتب النافذة من الطبع كتب الأطفال دليل الدوريات
شركة UMI	Business Periodical IEEE/IEE publication Social Science Index ABI / inform Dissertation Abetracts International	إدارة الأعمال الهندسة كشاف العلوم الاجتماعية إدارة الأعمال مستخلصات الأطروحات الجامعية

٢ / ٢ / ٣ نظم استرجاع تفاعلية ديناميكية:

تحتاج بعض نظم استرجاع الويب دون البعض الآخر لعنصر التفاعلية مع مستخدمي الويب. وتؤثر طبيعة محتوى الموقع في مدى احتياجه للتفاعلية أو ثبات المحتوى ؛ حيث تحتاج المواقع الإخبارية والإعلامية إلى تغيير محتوى الموقع بين يوم وآخر وذلك لما لطبيعة الأخبار اليومية من اختلاف. أيضًا فإن الدوريات العلمية ذات فترات الإصدار القصيرة تحتاج إلى تغيير أعدادها في كل إصدارة جديدة. على الجانب الآخر فإن نوع موقع الويب يؤثر أيضًا في احتياجه للتفاعل مع مستخدمي الويب ؛ حيث لا يمكن لمحركات البحث أو البوابات تنظيم وبحث المعلومات دون توجيه ومساعدة مستخدم الموقع بواسطة رسائل الحوار أو شاشات مساعدة البحث والبحث المتقدم.

ويقدم Robert Neches نموذجًا على نظم استرجاع الديناميكية في تغيير وتحديث المحتوى. يهتم هذا الموقع بمتابعة الأخبار الجغرافية على مستوى العالم؛ بالإضافة إلى أهم الأحداث المرتبطة بأماكن محددة وذات طابع ساخن، ومهمة هذا الموقع هو رصد الأخبار اليومية داخل مواقع الأخبار على الويب ، ثم تجميعها وإعادة تنظيمها مرة أخرى بما يتوافق ومعالجة الأماكن الجغرافية على مستوى العالم. ومن ثم يقدم موقع الويب المعلومات فور خروجها بما يحفظ له الحداثة الدائمة. وقد حدد النموذج مجموعة من المتطلبات للحفاظ على فاعلية وديناميكية مصدر الويب. وهي[١] :

(١) Ke-Thia Yao, Robert Neches. Dynamic Coordination of Information Management Services for Processing Dynamic Web Content, Information Sciences Institute, University of Southern California, 2005, cited 25/5/2007, cited at: http://www2002.org/CDROM/refereed/613

- **الشكل والاختيار**: حيث يتأثر محتوى موقع الويب بما يختاره من معلومات ومدى حداثة هذه المعلومات. إضافة إلى جودة خدمات المعلومات المقدمة.
- **إعادة تنظيم المعلومات**: يقدم موقع الويب التفاعلي مجموعة من المعلومات تم إصدارها بشكل منفصل، ويأتي دوره في مدى إعادة تنظيم وتشكيل هذه المعلومات بصورة مختلفة مرة أخرى.
- **تشكيل خدمات المعلومات** داخل مصدر الويب بما يوافق التغيير المستمر والتطوير في عرض المعلومات.
- **سلامة السياق العام** لعرض المعلومات: حيث قد يؤدي جمع معلومات مختلفة في إطار واحد إلى ضعف سياق عرض المحتوى العام.

٣ / ٢ / ٢ / ١ محركات البحث بوصفها نظمًا للاسترجاع على الويب:

تعددت التعريفات لمحركات البحث، ومن التعريفات التي وضعت لمحركات البحث : أنها "عبارة عن أداة تقوم بالبحث في و/ أو عن مصادر المعلومات على الإنترنت ـ والمصادر هنا يقصد بها المعلومات على المواقع وتخزين عناوينها على مرصد البيانات الخاص بها، ثم تقوم بإتاحتها للمستفيدين كل حسب المصطلح أو المصطلحات المستخدمة في البحث ومن ثم تمكن المستفيد من الوصول إلى مصادر المعلومات المختلفة على الإنترنت. ويتم تجميع هذه المصادر إما بطريقة آلية spidering or crawling أو بطريقة بشرية بواسطة الإنسان."(١)

---

(١) خالد محمد رياض. **أدلة ومحركات بحث شبكة الإنترنت : دراسة مقارنة**.- الإستراتيجية العربية الموحدة للمعلومات في عصر الإنترنت ودراسات أخرى .- تونس : الاتحاد العربي للمكتبات والمعلومات، (أكتوبر١٩٩٩م). ص ١٢٩.

وقد عرف زين عبد الهادي محركات البحث في البيئة العربية على أنها "أدوات بحث تعمل من خلال إستراتيجيات بحث محددة المنطق البوليني أو استراتيجيات بحث مفتوحة باللغة العربية، وذلك للبحث في حقول أو وثائق نصية، والأكثر من ذلك أنها تبحث عن أشياء كالمادة المصورة والخرائط والأشكال الأخرى في بيئة محددة هي شبكة الإنترنت وذلك يعني أنها تبحث في ملايين المواقع ومليارات الكلمات في وقت محدد وتتميز بسرعة الاستجابة وعادة ما تكون إجاباتها إما مواقع على الإنترنت تتوافر فيها كل المصطلحات التي تم البحث عنها أو بعضها، أو مواقع محددة سلفاً من خلال ما يعرف بأدلة البحث ".

لقد اعتمدت التعريفات السابقة لكل من خالد محمد رياض وزين عبدالهادي على التفصيلات المختلفة التي يتعرض لها محرك البحث في أداء مهامه على الويب، وعلى ذلك فقد اتخذت التعريفات السابقة الشكل الإجرائي للحديث عن محرك البحث .

و قد ذكرت بعض مواقع الويب تعريفات عن محركات البحث منها : " إن محرك البحث هو بصفة عامة أداة تمكن مستخدمي الويب من البحث بمصادرها، وأنه يمكن أن يعتمد على البرامج أو الإنسان في عمله[١]" وهذا التعريف قد عمل على تلخيص مهمة محرك البحث وإبرازها من ناحية مستخدمي الويب دون التطرق إلى الإجراءات التي يقوم بها في تنفيذ هذه المهمة. وثمة تعريف آخر هو "أنه عبارة عن أداة لإيجاد المعلومات على الإنترنت،

---

(١) Sullivan , Danny . Search Engine Optimization & Marketing Glossary, Submitawebsite, Inc , 2004 , cited at 12/6/2004 , cited at: http://www.submitawebsite.com/seo_glossary.html.

ويتكون من مكونات خمسة هي أولاً العنكبوت أو الروبوت، ثانيًا: المكشف، ثالثًا: قاعدة البيانات، رابعًا: برامج البحث، خامسًا: واجهة الويب[1]".
وقد اختلف هذا التعريف عما سبقه من ناحية التركيز على مكونات محرك البحث، وإن كان الباحث يرى أنها قد تقتصر فقط على ثلاثة مكونات هي الزاحف والمكشف والبحث.

وعلى ما سبق من تعريفات يرى الباحث أن أفضل تعريف لمحركات البحث هو " إحدى أدوات بحث استرجاع المعلومات على الويب، وتعتمد على ثلاثة برامج أساسية في ذلك هي الزاحف والمكشف وآليات البحث، بما يوفر القدرة على الإضافة والبحث. "

والجدير بالذكر أن هناك فرقًا بين مصطلحي محركات البحث وآليات البحث، فالأول إنما يعني برامج لبحث مصادر معلومات الويب، أما الثاني فهو طرق البحث داخل قواعد بيانات محركات البحث مثل آليات المنطق البوليني، والبتر الضمني، واللغة الطبيعية، والبحث بالمعنى أو المفهوم... إلخ. وعلى ذلك فإن ثمة وجهين تعمل بهما محركات البحث؛ الوجه الأول وهو الذي تتعامل به مع الإنترنت في عملية التجميع والفهرسة والتكشيف لمواقع الويب والوجه الثاني الذي تتعامل به مع مستخدمي الويب طلبًا للمعلومات. والوجه الثاني قد يكون مشرقًا أمام بعض الباحثين بالحصول على المعلومات التي يريدونها بسهولة ويسر، وقد يكون عبوسًا أمام البعض الآخر لما يلاقيه هؤلاء من نتائج غير مطابقة لموضوعاتهم وأسئلتهم البحثية، تنتج عن استخدام المصطلحات غير المناسبة لاستدعاء مصادر الموضوع.

---

(١) What is a Search Engine? Metamend.inc , 2004 , cited at 12/6/2004 , cited at http://www.metamend.com/search-engine-map.html.

فالأمر هنا ليس مجرد الدراية بآليات البحث والتمرس بها ، بل يتعداه إلى حد وعي الباحث بجوانب الموضوع؛ فعلى سبيل المثال فإن سؤالا مثل (ما هو التدخين ؟) فإنه ينطوي على جانب موضوعي واحد هو (التدخين)، وسؤالا عن (هل يسبب التدخين السرطان ؟) فإنه ينطوي على جانبين هما : (التدخين، والسرطان)، كما إن سؤالا عن (هل يعلم الأطفال أن التدخين يسبب السرطان ؟) يتشابك فيه مجموعة الجوانب " الأطفال والتدخين والسرطان". وعلى ذلك فلا بد أن يكون الباحث على دراية بماهية الموضوع وجوانبه المختلفة لكي يستطيع تحليله ووضع المصطلحات التي تقيه أخطاء وتشويش النتائج في محركات البحث التي لا تعرف سوى مطابقة أو مضاهاة كلمات النص جميعها مع مصطلحات البحث.

ومن ثم فيمكن أن نلخص الأسباب التي تدفع الباحثين لاستخدام محركات البحث في:

١- أن يكون لدى الباحث فكرة أو موضوع مخصص داخل ذهن الباحث.

٢- أن يهدف الباحث الوصول إلى موقع محدد أو مخصص.

٣- أن يهدف الباحث بحث النص الكامل لملايين من صفحات ومواقع الويب.

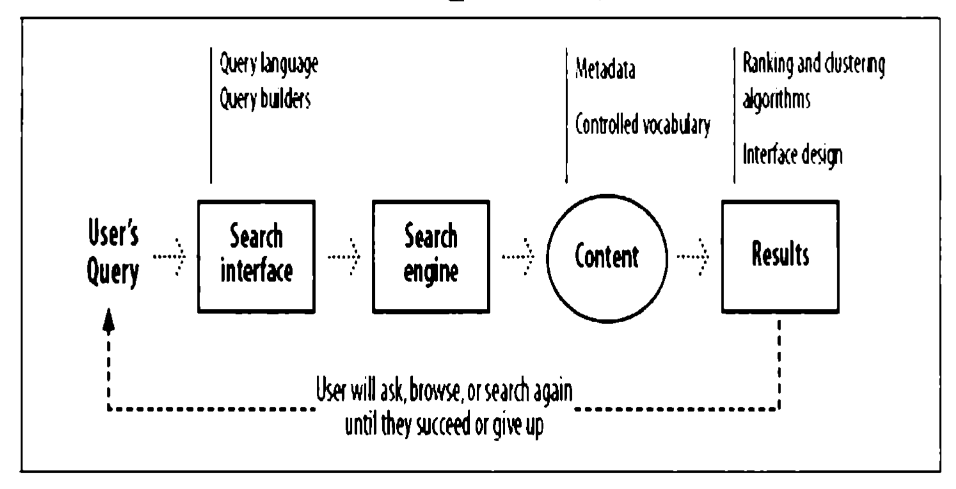
٤- أن يهدف الباحث استدعاء كم كبير من مواقع وصفحات الويب.

٥- أن يهدف الباحث بحث نوع معين من صفحات أو ملفات الويب أو مواقع للغة معينة.

٦- أن يرغب الباحث في الحصول على ميزات تدعم بحث الويب مثل وزن النتائج المسترجعة أو البحث بالمفاهيم.

٣ / ٢ / ٢ / ٢ التفاعلية في محركات البحث:

تعمل محركات البحث كنظم لاسترجاع معلومات الويب بهدف تقديم مختلف مصادر المعلومات الرقمية التي يحتاجها المستفيدون بحسب احتياجاتهم الموضوعية. ومن ثم فإن هناك مجموعة من المراحل التي يمر بها التفاعل بين النظام والمستفيد ، وهو بطبيعته يختل عن التفاعل بين النظام ووثائق الويب كمصادر للمعلومات. ويعبر الشكل عن مراحل التحاور والتفاعل التي تبدأ بطرح استفسار البحث وفقاً للغة نظام الاسترجاع، ثم تلقي نتائج البحث بعد قيام نظم محركات البحث بدورها في ترتيب وتنظيم قائمة النتائج النهائية للمستفيد.



الشكل رقم (٣-٧) يوضح مراحل التفاعل بين محرك البحث كنظام للاسترجاع ومستخدمي الويب

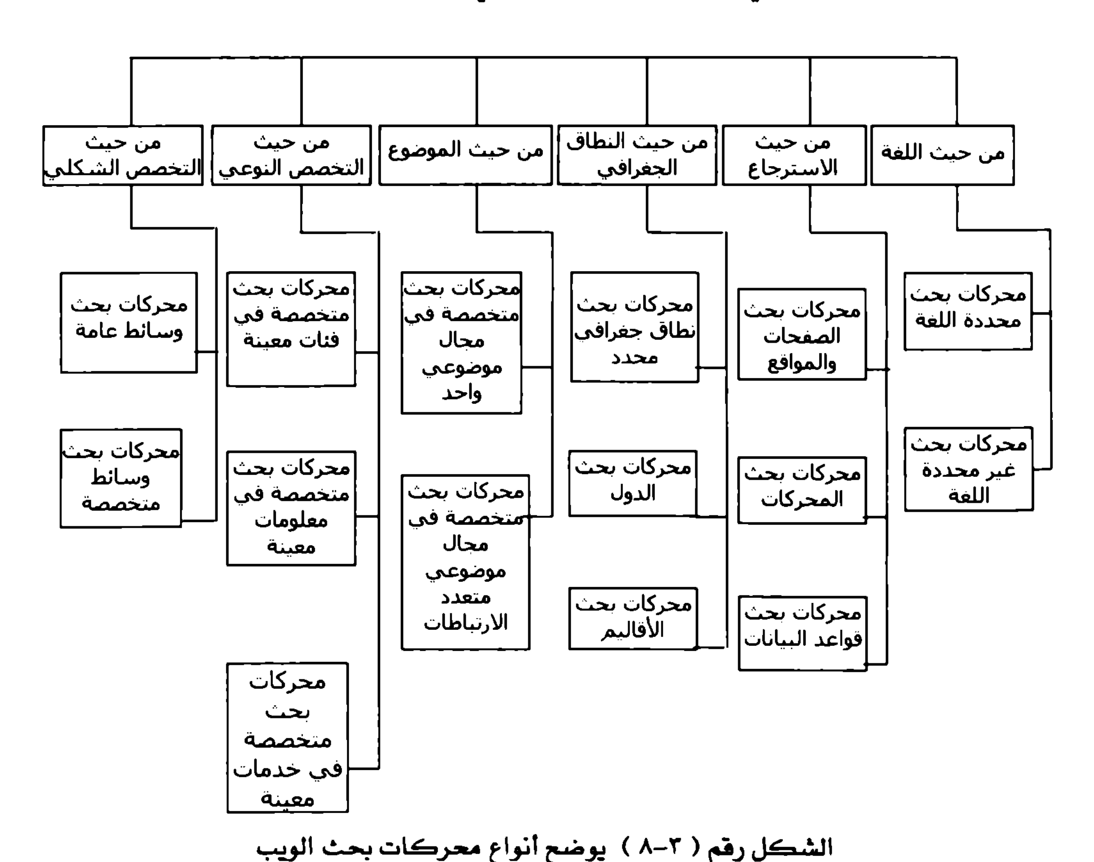
٣ / ٢ / ٢ / ٣ أنواع محركات البحث:

تعتمد محركات البحث على نوع واحد من البرامج المستخدمة والتي يتشابه أداؤها في كل أنواع المحركات وسوف يكون التقسيم هنا ليس لإبراز هذه الأنواع وإنما لبيان الجوانب التي يمكن أن تتناول أنواع محركات البحث

عن طريقها. ويوضح الشكل رقم (١-٨) أنواع محركات البحث مقسمة وفق التخصص.وفيما يلي شرح لهذه الأنواع:

٣ / ٢ / ٢ / ٢ / ٣ / ١ أنواع المحركات من حيث اللغة:

تستخدم الغالبية العظمى من المحركات اللغة الإنجليزية بالإضافة إلى بعض اللغات الطبيعية الأخرى في البحث عن الصفحات والمواقع المتاحة على الويب، إلا أن هناك عددًا من المحركات تسمح باختيار لغة النص الذي يبحث فيه، واستبعاد ما دونه من اللغات. ويمكن بذلك تقسيم محركات البحث حسب لغات النص التي تبحث فيها إلى ما يلي:



الشكل رقم (٣-٨) يوضح أنواع محركات بحث الويب

١- **محركات بحث غير محددة اللغة**: حيث لا يعطي محرك البحث هنا خيارًا للغة النص، وإنما يقوم محرك البحث هنا بمضاهاة الكلمات التي جاءت في إستراتيجية البحث في أي صفحات وجدت. ويفيد مثل هذا النوع من المحركات الباحثين الذين لا يسعون إلى تحقيق عالٍ وإنما يأتي الاستدعاء في المقام الأول. هذا بالإضافة إلى أن لغات النشر الإلكتروني لم تعد تقتصر على مجموعة لغات معينة وإنما أصبح كل من يمتلك حاسبًا آليًا ومكانًا على أحد خادمات الإنترنت باستطاعته النشر وإلقاء المعلومات التي يريدها في فضاء الويب. ومن ثم أصبح من غير الممكن أن يلم الباحث بكل اللغات التي ينشر بها الإنتاج الفكري في موضوع بحثه. ومن نماذج محركات البحث غير محددة اللغة محرك البحث Altavista (١) ومحرك البحث Google (٢) ومحرك البحث Alltheweb(٣).

٢- **محركات بحث محددة اللغة**: يأتي هذا النوع عكس سابقه، حيث يسمح المحرك بالتحديد في البحث بلغات النصوص المنشورة بها صفحات الويب بطرق مختلفة منها؛ الطريقة الأولى : أن يقوم محرك البحث باستدعاء كل الصفحات التي تحوي الكلمات المفتاحية ضمن نصوصها بغض النظر عن اللغة، وإذا ما أراد الباحث تحديد لغة النص فإن محرك البحث يمده بقائمة من اللغات الطبيعية المختلفة

---

(١) http://www.Altavista.com.

(٢) http://www.Google.com.

(٣) http://www.Alltheweb.com.

يختار منها الباحث ما يناسبه فيتم استبعاد الصفحات التي لا تقع مفردات نصها ضمن اللغة المختارة، وهنا تدخل اللغة كعامل محدد إضافي لنوع النتائج والوحدات التي يخرجها محرك البحث. الطريقة الثانية : أن يفرض محرك البحث مسبقًا لغة النص التي يرغب البحث عنها، وهذه ترجع إلى الحالات التي يكون فيها محرك البحث مختصًا بمنطقة محددة تعتمد لها لغة خاصة بها.

ومن نماذج محركات البحث محددة اللغة:

- محرك البحث Google [١]
- محرك البحث باللغة الإسبانية[٢]
- محرك البحث باللغة الفرنسية[٣]

٣ / ٢ / ٢ / ٣ / ٢ أنواع المحركات من حيث الاسترجاع:

يقصد بأساليب الاسترجاع هنا هو كيف يتم استرجاع مواقع وصفحات الويب هل بالأسلوب المباشر أي باللجوء إلى بيانات الفهرسة للصفحات والمواقع، أم بالأسلوب غير المباشر أي من خلال أدوات بحث أخرى. وتنقسم محركات البحث من حيث أساليب الاسترجاع إلى ثلاثة أنواع:

١- **محركات بحث الصفحات والمواقع** : وهي الأكثر انتشارًا وتعد بمثابة محركات بحث من الدرجة الأولى وتعمل على جمع وتنظيم وبحث

---

(١) http://www.Google.com.

(٢) http://www.sportec.com/.

(٣) http://www.pratique.fr/.

مواقع وصفحات الويب بشكل مباشر من خلال الاطلاع على بيانات الفهرسة الخاصة بها، مثل URL الخاص بالصفحة مع مجموعة الواصفات التي تحدد الهوية الموضوعية لهذا الموقع. أو من خلال التعرف إلى كلمات وردت في النص.

ومن نماذج محركات بحث الصفحات والمواقع:

- محرك البحث Google (١)
- محرك البحث Yahoo (٢)
- محرك البحث Altavista (٣)

٢- **محركات بحث المحركات** : ويعرف هذا النوع بالإنجليزية باسم Meta Search Engines. وهي تعد بمثابة محركات من الدرجة الثانية، وهي تبحث في نتائج البحث التي تأتي بها محركات بحث الدرجة الأولى. وهذا النوع يشبه مجازًا ما تكون عليه ببليوجرافيات الببليوجرافيات، إلا أن بعض محركات بحث المحركات تتعدى الحصر لنتائج محركات الدرجة الأولى إلى البحث والتنقيب المباشر في صفحات الويب. وتعرض هذه المحركات نتائج البحث إما بعرض نتائج البحث حسب كل محرك أو أن يتم عرضها بصفة عامة دون التحديد.

ومن نماذج محركات بحث المحركات:

---

(١) http://www.Google.com.

(٢) http://www.yahoo.com.

(٣) http://www.Altavista.com.

- محرك البحث Alltheweb [١]
- محرك البحث MetaCrawler [٢]
- محرك البحث Mamma [٣]

٣- **محركات بحث قواعد البيانات** : يعمل هذا النوع على اكتشاف الجزء غير المرئي من الويب، حيث مازال يطلق على قواعد البيانات مصطلح the invisible web. وتحاول هذه المحركات إخضاع محتويات قواعد البيانات من المقالات ومصادر المعلومات للبحث أمام الباحثين دون اللجوء إلى مواقع تلك القواعد، ذلك من أجل إتاحة كل مصادر معلومات الويب لإفادة الباحثين. ولقد وقفت طبيعة بناء قواعد البيانات حائلا أمام قدرة ببرامج الزاحف لاقتناء وإضافة الصفحات المتعمقة داخل هذه القواعد، كما تسعى محركات البحث على رفع قدرة برامج الإضافة بها لتحليل صفحات تلك القواعد بما يمكن معه اقتناء جميع وحدات المعلومات داخل قاعدة البيانات .

ومن نماذج محركات بحث قواعد البيانات

- محرك البحث Invisibleweb [٤]
- محرك البحث Webdata [٥]

---

(١) http://www.Alltheweb.com.

(٢) http://www.metaCrawler.com.

(٣) http://www.mamma.com.

(٤) http://www.Invisibleweb.com.

(٥) http://www.Webdata.com.

٣ / ٢ / ٢ / ٣ / ٣ أنواع المحركات من حيث النطاق الجغرافي :

يتمثل التخصص الجغرافي في أن يحصر المحرك بحثه في المواقع التي تقع أو تتبع حدود جغرافية محددة؛ حيث يقوم المحرك بجمع وتنظيم المواقع التي تخص منطقة أو إقليم جغرافي محدد. وعادة يلجأ في هذه الحالة إلى البحث عن النطاق المعرف للدولة. وهذا النوع يوجد منه الكثير على الويب، ويمكن تقسيم محركات البحث من حيث التخصص الجغرافي إلى ما يلي:

١- **محركات بحث ذات نطاق جغرافي محدد**: أي أن محركات بحث هذا النوع يمكن أن تفيد الباحثين بتقديم مواقع وصفحات الويب التي تنتمي إلى قارة من القارات، مثل المحركات التي تختص بمواقع القارة الأوربية أو تلك المحركات التي تهتم بمواقع وصفحات الويب التي تنتمي إلى قارة أمريكا الشمالية. وقد تتداخل هذه المحركات مع تلك الأنواع المتخصصة في اللغة وهي المناطق ذات اللغات المحدودة جغرافيًا.

ومن نماذج محركات البحث ذات النطاق الجغرافي المحدد:

- محرك بحث عن آسيا Lycos Asia
- محرك بحث عن كندا canadaspace (١)
- محرك بحث عن ولاية new Mexico (٢)

---

(١) http://www.canadaspace.com.

(٢) http://www.nmmagazine.com/cgi-bin/links.cgi.

٢- **محركات بحث الدول** : تقتصر عملية الإضافة والبحث داخل هذا النوع فقط على مواقع وصفحات الويب المنتمية إلى الدولة التي يعمل المحرك في نطاقها. وذلك من خلال محددات في تصميم مواقع الويب يمكن من خلالها الاستدلال على الدولة المنتجة للموقع. وقد يتعدى الأمر إلى المعنى الذي تقوم عليه الببليوجرافيا الوطنية؛ حيث تجمع ما ينتمي إلى الدولة بالمعنى الواسع مثل الذي تنشره الدولة، وما ينشر عنها، وما تنشره مؤسسات تابعة لها في أي موقع في العالم.

ومن نماذج محركات بحث المناطق والدول :

- محرك البحث عن مصر Egyptsearch [١]
- محرك البحث عن اليابان SearchDesk [٢]
- محرك البحث عن فرنسا Abondance [٣]

٣- **محركات بحث الأقاليم** : تعمل محركات بحث الأقاليم على توسيع نطاق بحث المواقع بشكل أكبر اتساعًا من محركات بحث الدول، وفي الوقت ذاته لا تكون على نفس قدر الاتساع في المواقع الشاملة. ويأتي هذا النوع كنتيجة لتشابه مجموعة من الدول في اللغة والنواحي البحثية أو الاجتماعية. ويمكن أن يقال الثقافة في هذه الدول مثل الشرق الأوسط، أو الخليج العربي، أو دول شمال أوربا المعروفة باسم

(١) http://www.egyptsearch.com/search/.
(٢) http://www.searchdesk.com/.
(٣) http://webmasters.abondance.com/.

اسكندنافيا، أو دول جنوب شرق آسيا والصين وتايوان والهند وهونج كونج. وتوفر هذه المحركات الدرجة العالية من الاستدعاء لمواقع موضوع البحث مع وجود حدود يمكن الرجوع إليها للتحكم في النتائج. وفيما يلي أمثلة لهذا النوع:

ومن نماذج محركات بحث الأقاليم:

- محرك البحث Almashriq (Middle East) (١)
- محرك البحث Latin America (٢)
- محرك البحث SearchEurope (٣)

٣ / ٢ / ٢ / ٣ / ٤ أنواع المحركات من حيث الموضوع:

هناك الكثير من محركات البحث على الويب تقتصر في بحثها على تخصص موضوعي واحد مثل الطب، أو قد يتسع ذلك ليغطي عددًا من الموضوعات ذات الصلة كالطب والصحة والجراحة. وتأخذ هذه المحركات في الغالب شكل قواعد البيانات على الويب. وهذا النوع من قواعد البيانات يهتم بموضوع محدد ويجمع كل ما يمكن من مواقع ومصادر معلومات إلكترونية تفيد الباحثين والمهتمين بهذا الموضوع. ومن ثم فإن عمل محرك البحث هنا لا يتفق كثيرًا مع المعنى الذي يحمله مصطلح محرك البحث. إلا أن الباحث قد لاحظ ذكر أنواع قواعد البيانات في تصنيف محركات البحث داخل الفئة

---

(١) http://almashriq.hiof.no.

(٢) http://www.yupimsn.com/index.asp.

(٣) http://www.searcheurope.com.

المتخصصة موضوعيًا. وتنقسم محركات البحث من حيث الموضوع إلى نوعين رئيسيين هما:

١- **محركات بحث متخصصة في مجال موضوعي واحد** : وهي المحركات التي تهتم بالبحث في أحد الموضوعات دون النظر إلى بقية فروع المعرفة، ومن أهم الأمثلة على هذا النوع ما يلي:

ومن نماذج محركات البحث المتخصصة في موضوع واحد:

- محرك بحث متخصص في الطب Midisearch [١]
- محرك بحث متخصص في القانون Lawcrawler [٢]
- محرك بحث متخصص في العلوم Search4Science [٣]

٢- **محركات بحث متخصصة في مجال موضوعي متعدد الارتباطات:** وهي المحركات التي تعمل في مجال موضوعي يغطي مجالات معرفية ذات علاقة ببعضها البعض كالعلوم الاجتماعية.

ومن نماذج محركات البحث المتخصصة في موضوع متعدد الارتباطات:

- محرك البحث في العلوم الصحية mwsearch [٤]
- محرك البحث في الإنسانيات eserver [٥]

---

(١) http://midisearch.net/.

(٢) http://findlaw.com.

(٣) http://www.search4science.com/.

(٤) http://www.mwsearch.com/.

(٥) http://www.eserver.org/.

- محرك البحث في إدارة الأعمال business [١]

٣ / ٢ / ٢ / ٣ / ٥ أنواع المحركات من حيث التخصص النوعي:

وتتخصص هذه المحركات في تغطية صفحات ومواقع الويب التي تهم فئة معينة من الباحثين أو تغطي معلومات، أو خدمات معينة. وتنقسم محركات البحث هذه إلى ثلاثة أنواع رئيسة هي:

١- **محركات بحث متخصصة في فئات معينة من المستفيدين** : ومن أمثلة هذا النوع محركات البحث المتخصصة في مواقع الأطفال.

ومن نماذج محركات بحث الأطفال:

- محرك البحث KidsClick [٢]
- محرك البحث Ask Jeeves for kids [٣]
- محرك البحث Yahooligans [٤]

٢- **محركات بحث متخصصة في معلومات معينة**: ومن أمثلة هذا النوع محركات البحث المتخصصة في الأخبار.

ومن نماذج محركات بحث الأخبار:

- محرك البحث Google News [٥]

---

(١) http://www.business.com/.

(٢) http://www.kidsClick.com.

(٣) http://www.ajkids.com.

(٤) http://www.Yahooligans.com.

(٥) http://news.Google.com/.

- محرك البحث Yahoo News [١]
- محرك البحث Altavista News [٢]

٣- **محركات بحث متخصصة في خدمات معينة**: ومن أمثلة هذا النوع محركات البحث المتخصصة في خدمات التسوق.

ومن نماذج محركات بحث التسوق :

- محرك البحث Froogle [٣]
- محرك البحث MSN Shopping [٤]
- محرك البحث Yahoo Shopping [٥]

٣ / ٢ / ٢ / ٣ / ٦ أنواع المحركات من حيث التخصص الشكلي:

تبحث هذه المحركات في أشكال معينة من وسائط المعلومات مثل الصوت أو المادة المصورة أو لقطات الفيديو. وهي تعتمد في بحثها لهذه الملفات على رموز لغة النص الفائق html ورموز meta tag ، وتستطيع بذلك التعرف إلى URL وعمل رابط له داخل قواعد بياناتها. وهي تنقسم إلى نوعين من المحركات هي:

---

(١) http://news.yahoo.com/.

(٢) http://news.Altavista.com/.

(٣) http://wwwfroogle.com.

(٤) http://www.msnshopping.com.

(٥) http://www.yahooshopping.com.

١- **محركات بحث عامة** : ويطلق على مثل هذا النوع من محركات البحث مصطلح General Media Search Engines التي تجمع مختلف أنواع الوسائط التي تحمل المعلومات.

ومن نماذج محركات بحث الوسائط المتعددة :

- محرك البحث Alltheweb (١)
- محرك البحث Altavista(٢)

٢- **محركات بحث متخصصة** : وهي محركات متخصصة في شكل معين من الوسائط مثل الفيديو، حيث تقوم باقتناء وتكشيف وبحث ملفات الفيديو من خلال سمات ملف الفيديو الإلكترونية Content Based Video Retrieval. ثم تعمل على الترتيب الإلكتروني لملف الفيديو كمقاطع مع ترتيب مشاهد الفيلم التي تدور حول مكان أو زمان واحد، كما يتم ترتيب لقطات كل مشهد أيضًا ثم وصف كل وحدة من هذه الوحدات بحسب بداية الفيديو إلى نهايته، هذا بجانب الوصف الببليوجرافي لملف الفيديو. وكذلك محركات بحث المواد المسموعة، حيث تعتمد على الروابط التي تضعها مع أسماء الأغاني أو ما شابهها من الملفات الأخرى . كما يغلب على هذا النوع العمل بأسلوب التصفح من قوائم تبدأ عادة بالمغني ثم الأعمال التي أنتجها، ومن ثم يعتمد الاسترجاع على ربط الأغاني بأسمائها. وكذلك محركات بحث المواد المصورة وهي كثيرة وسيتم تناولها بالتفصيل في هذا الفصل.

---

(١) http://multimedia.Alltheweb.com.

(٢) http://www.Altavista.com/mutimedia/default?qbmode=.

ومن نماذج محركات البحث المتخصصة:

- محرك بحث صوت فقط The Music Finder (١)
- محرك بحث صور فقط Picsearch (٢)
- محرك بحث صور وصوت فقط Lycos Pictures & Sounds(٣)
- محرك بحث صوت وفيديو فقط Singingfish (٤)

٣ / ٣ بزوغ الويب الدلالية:

كان لزامًا على الباحث هنا أن يتعرض إلى تقنية الويب الدلالية semantic web لما تعكسه الآن من تقدم في خصائص التفاعلية التي تسير عليها الويب. فالويب مع نظم الاسترجاع التي نعرفها عملت على معالجة واسترجاع المعلومات الرقمية دون أن تفهم دلالة أو معنى هذه المعلومات، من ثم فإن الكثير من جوانب القصور قد أصابت محاولات الويب الرامية لتحسين أداء نظم استرجاعها. أما الويب الدلالية فتعمل على تحقيق الآتي:

أولاً: جعل نظم استرجاع معلومات الويب والحاسبات الآلية التي تستخدمها، قادرة على فهم وإدراك دلالة المصطلحات البحثية التي يستخدمها المستفيد للتعبير عن احتياجاتها الموضوعية.

---

(١) http://www.music-finder.net/.

(٢) www.Picsearch.om.

(٣) http://multimedia.lycos.com/.

(٤) http://search.singingfish.com/sfw/home.jsp.

ثانيًا : إذا ما أدركت نظم الاسترجاع دلالة الكلمات البحثية والمصطلحات المستخدمة، فإنها تستطيع مساعدة المستفيد بمعاودة البحث عن مختلف المعلومات التي توافق هذه الدلالة الموضوعية. أي أن الويب ونظم استرجاعها الآن قد أقبلت على مرحلة المضاهاة والمقابلة بالفهم والإدراك وليس بالشكل والصورة كما اعتادت نظم المضاهاة.

لقد ذكر رجب عبد الحميد أن هذه التقنية قد نمت بذورها في عقل Tim Berners-Lee أحد القائمين على دعم وتطوير بيئة الويب. كما أن مفهوم الويب الدلالي ينطوي على أنها شبكة بيانات بالمعنى ؛ أي أن البرامج و الحاسبات الآلية تستطيع فهم ومعرفة معنى البيانات المتداولة عليها، وذلك بحسب ما أوردت الشركة القائمة على الويب W3C أو رابطة الشبكة العالمية (World Wide Web Consortium). ويتم اعتماد هذه التقنية باستخدام اللغات المتقدمة في التعبير عن البيانات على لغات البرمجة المتقدمة مثل XML (Extensible Markup Language) أو لغة الترميز الموسعة، وخرائط المفاهيم أو الوجوديات Ontology، والمعيار العام لوصف المصادرschema RDF (Resource Describe Framework)، ولغة الويب الوجودية OWL (Ontology Web Language). وقد أشار Sean B. Palmer, إلى أنه من خلال الويب الدلالي نستطيع استخدام مختلف المعلومات التي تنشر على الويب دون إمكانية الاستفادة منها، وهذه المعلومات مثل نشرات الأرصاد، وبيانات التوقيت وإحصاءات الألعاب الرياضية المختلفة. كل ذلك تستطيع الويب الدلالية استخدامه وعرضه بشكل ثلاثي الوجهة والرجوع إليه لتحقيق الإفادة التامة منه، أي أنه يمكن النظر إلى

الويب الدلالي على أنه مهندس الصيانة لمختلف المعلومات الرقمية غير القابلة للمعالجة كملفات ذات بنية مختلفة عن بنية صفحات الـ html. [١][٢]

٣ / ٤ الخاتمة:

لقد مرت نظم استرجاع المعلومات من التقليدية إلى الويب الدلالية بقدرات ومستويات مختلفة من التفاعلية والقدرات الديناميكية التي تعمل بها. فمنذ بدايتها تعمل نظم استرجاع المعلومات على تطوير قدراتها في تحقيق التفاعل والكفاية في استرجاع المعلومات المحددة التي يحتاجها المستفيدون، دون التعرض للزيادة المضللة أو الفقد المخفق في المعلومات. وقد عكس محتوى هذا الفصل طبيعة محتوى الويب ونظم استرجاعها التي اتسمت بالتغيير المتلاحق والتسارع في أنشطتها ؛ حيث اتضح ذلك جليًا في ديناميكية العمل داخل أدوات بحث الويب، مرورًا بالتعرض لأهم نماذج الويب التفاعلية كمحرك البحث ، ثم انتهاءً بالويب الدلالية كأحدث تقنيات التفاعل بين الويب بوصفه نظامًا كونيًا لاسترجاع المعلومات وبين المستفيد.

---

(١) رجب عبد الحميد .**تقنيات الويب الدلالي للمكتبات الرقمية** -. cybrarians journal -. ع ١٤ (سبتمبر ٢٠٠٧) .- تاريخ الإتاحة > اكتب هنا تاريخ اطلاعك على الصفحة - . < متاح في :
http://www.cybrarians.info/journal/no14/semantic.htm

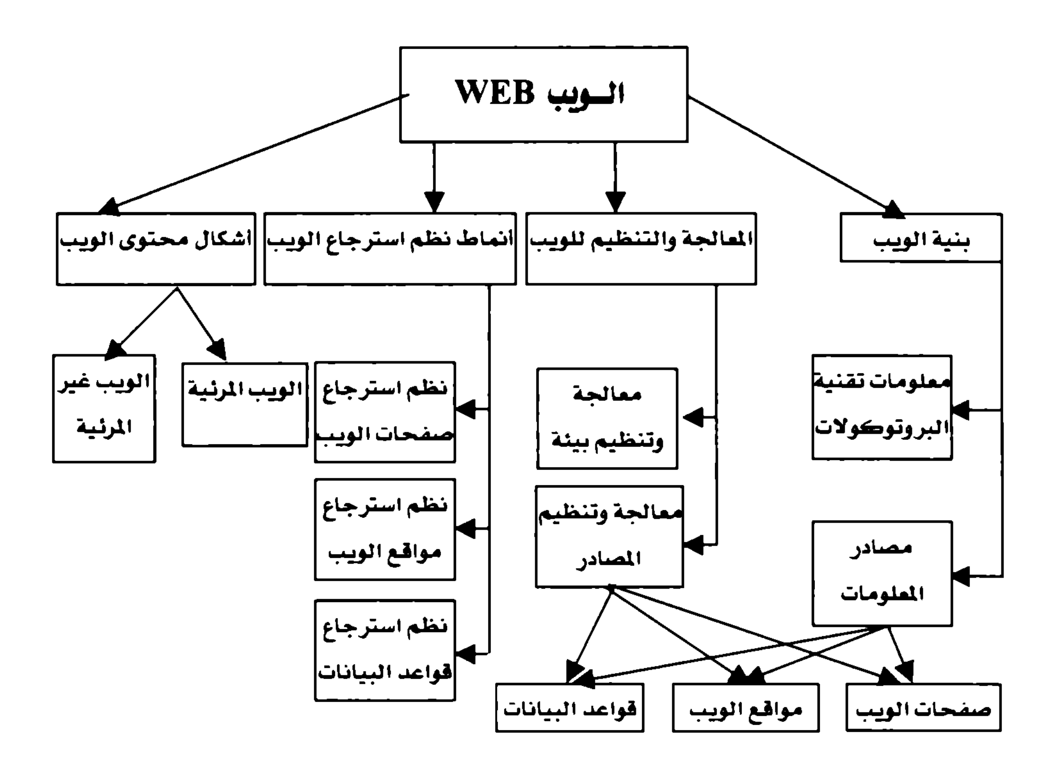
(٢) Palmer, Sean B. The Semantic Web: An Introduction, infomeshnet, 2001, cited 5/1/2008, cited at http://infomesh.net/2001/swintro/.

الفصل الرابع

الويب نظام استرجاع كوني

٤ /٠ التمهيد:

لقد كونت الويب عالمًا من المعلومات الرقمية متعدد الموضوعات مختلف الأشكال، يختلف هذا العالم الرقمي عن نظيره التقليدي في التعامل مع المعلومات معالجة وتنظيمًا واسترجاعًا. وتطورت أشكال مصادر معلومات الويب منذ بدايتها لتمر بالأشكال النصية والسمعية والمرئية، وصولاً إلى ما يعرف بمراصد أو قواعد البيانات. وانقسمت الويب منذ ظهور قواعد البيانات إلى شقين ؛ الأول: مرئي متمثلاً في صفحات ومواقع الويب، والثاني غير مرئي أو عميق أو مخفي متمثلاً في ملفات قواعد البيانات. وتعمل الدراسة في هذا الجزء على دراسة ماهية الويب عامة والويب غير المرئية خاصة، ثم تتابع الفصول للتعرف إلى خصائص بيئة الويب غير المرئية وسبل التعامل الفني من تنظيم واسترجاع بما يتناسب مع طبيعة بناء مصادر قواعد البيانات. وتأتي الأهمية الكبرى للدراسة في محاولة إيجاد الشكل الملائم من تعامل محركات البحث مع قواعد البيانات إضافة وتنظيمًا وبحثًا. غير أن رؤيتنا للويب سوف تختلف عن نظرة الدراسات الأخرى للويب ؛ حيث تركز الدراسة هنا على الجوانب التنظيمية والهيكلية والعوامل التي تقف خلف ظهور الويب بما نشاهدها عليه عبر نوافذ استخدام الإنترنت. وتأتي عناصر هذا الفصل لتنصب على بيئة الويب بنية وتنظيمًا ومصادر، وصولاً إلى تناول أحد مصادر الويب وهي قواعد البيانات التي تشكل بمفردها بيئة جديدة تسمى بيئة الويب غير المرئية. ويوضح الشكل رقم (٤ / ١) ملامح الويب:



الشكل رقم (٤ / ١) بنية وتنظيم الويب وبيئة الويب غير المرئية.

٤ / ١ بنية الويب :

قبل التعريف ببنية الويب ، يمكن وصف الويب على أنها " شبكة عالمية تقدم خدمات الإنترنت ، وتجهيز الوثائق التي يتم بناؤها بلغة ترميز النص الفائق التي تسمح باسترجاع الوثائق باستخدام الروابط الفائقة ، وتم تصميم الويب في عام ١٩٨٩م عن طريق طالب في جامعة الينوى University of Illinois وارتبطت بداية الويب باستخدام متصفح Mosaic ثم خروج متصفح نيتسكيب Netscape Web browser " وذلك وفق ما جاء في قاموس مصطلحات المكتبات والمعلومات على الخط المباشر ODLIS تحت مسمى الشبكة العالمية

العنكبوتية World Wide Web (WWW). ويتفق الباحث مع تعبير المفهوم على أن الويب باعتبارها شبكة عالمية إنما هي جزء من الإنترنت تقدم خدمات ذات طابع يتفق مع بنية وتنظيم الويب شبكة الويب[1] .

وتعتمد الويب بوصفها بيئة رقمية في بنيتها على شقي بنية النظام الرقمي ؛ وهما الجانب المادي hard ware من التجهيزات والجانب الفني من البرامج soft ware. غير أن الحديث هنا سيركز فقط على الجانب الفني من بنية الويب وهي البرامج والمعلومات النظم التي تعمل بها الويب. وتحتوي الويب على عناصر معلوماتية تختلف بين بروتوكولات عمل الويب وتنظيم المعلومات حفظًا واستدعاء، بالإضافة إلى المعلومات المسجلة الموزعة على أشكال وموضوعات مختلفة. وتم توضيحها كما يلي :

٤ / ١/ ١ المعلومات الفنية التقنية(بروتوكولات الويب):

تتنوع بروتوكولات عمل الويب نظرًا لتنوع الخدمات والمعلومات التي تقدمها. غير أن أشهر هذه البروتوكولات هو بروتوكول نقل النص الفائق HTTP، وبروتوكول نقل معلومات الإنترنت IP، وبروتوكول البريد الإلكتروني SMTP. ويوضح الجدول (٤ - ١) مختلف البروتوكولات التي تتعامل بها الويب [2]

(١) ODLIS —Online Dictionary for Library and Information Science. Cited 10/5/2007, cited at http://lu.com/odlis/odlis_w.cfm.

(٢) Web protocols. Securepaynet, 2006, cited 10/5/2006, Cited at http://www.realifewebdesigns.com/web-resources/web-protocols.html

الجدول رقم (٤ / ١) بروتوكولات عمل الويب

Protocol	Description
ARP	Address Resolution Protocol
HTTP	Hypertext Transfer Protocol
FTP	File Transfer Protocol
SMTP	Simple Mail Transfer Protocol
DNS	Domain Name Service
DSN	Data Source Name
IP	Internet Protocol
TELNET	TCP/IP Terminal Emulation Protocol
IMAP	Internet Message Access Protocol
ICMP	Internet Control Message Protocol
IDRP	ICMP Router-Discovery Protocol
DHCP	Dynamic Host Configuration Protocol
IRC	Internet Relay Chat Protocol
POP3	Post Office Protocol version 3
PAR	Positive Acknowledgment and Retransmission
RLOGIN	Remote Login
SSL	Secure Sockets Layer
SSH	Secure Shell
TCP	Transmission Control Protocol
UPD	User Data gram Protocol
UPS	Uninterruptible Power Supply

يعد بروتوكول http أهم البروتوكولات السابقة نظرًا لما له من ارتباط وثيق بآلية عمل الويب؛ حيث يعد التصفح والإبحار داخل الويب web navigation هو أهم أساليب استرجاع المعلومات على الويب. يلي ذلك بروتوكولات خدمات الويب الأخرى مثل البريد الإلكتروني SMTP ونقل الملفات FTP.

٤ / ١/ ٢ مصادر معلومات الويب :

يقصد بمصادر معلومات الويب هنا مختلف أشكال حفظ المعلومات المراد إتاحتها عبر الويب بما تحويه من وسائط مختلفة لحمل المعلومات مثل النص والصوت والصورة. وقد اقترح الباحث تقسيم مصادر المعلومات بحسب إمكانات حفظ وعرض المعلومات بدلاً من التقسيم بحسب أشكال المعلومات ذاتها ، وذلك اتساقًا مع الحديث عن تنظيم المعلومات داخل بنية الويب.

وقد استثنى الباحث من هذه المصادر الحديث عن الملفات الرقمية التي تنتمي إلى صفحات ومواقع الويب ذلك لما لهذه الملفات من خصائص تختلف بها عن منظومة بنية الويب. ومن بين هذه الخصائص الآتي:

- تعتبر الملفات الرقمية وحدات لا يمكن تقسيمها عكس صفحة الويب التي تعد وحدة بناء للويب، وتحتوي هذه الوحدة على ملفات تحمل معلومات متعددة الوسائط.
- إن ملفات الويب لا يمكن تواجدها على الويب بمفردها دون صفحة مضيفة host web page لهذه الملفات تحويها في إطار عرض الموضوع العام.
- إن ملفات الويب لا تحمل خصائص شكلية تؤثر في تنظيم الويب، وإنما تحمل الملفات الرقمية خصائص ما تحمله من معلومات مثل خصائص النص للملفات النصية أو خصائص المعلومات المرئية لملفات الصور الثابتة والمتحركة أو خصائص المعلومات السمعية لملفات الصوت.

- تحتاج الملفات الرقمية إلى المعالجة الفنية بمعايير معالجة وسائط حمل المعلومات" النص والصورة والصوت"، بينما تحتاج صفحات ومواقع وقواعد البيانات إلى معايير معالجة مصادر المعلومات المكونة من خصائص شكلية وموضوعية.

٤ / ١/ ٢ / ١ صفحات الويب:

لقد عرف قاموس مصطلحات المكتبات والمعلومات ODLIS صفحة الويب على أنها " وثيقة إلكترونية تمت كتابتها بلغة ترميز النص الفائق HTML ويتم استرجاعها باستخدام متصفح الويب عن طريق محدد المصادر الموحد URL، وتمثل صفحة الويب في أحيان كثيرة تجمعًا لمجموعة روابط ملفات إلكترونية ذات أنواع مختلفة، ويمكن لصفحة الويب أيضًا احتواء أشكال مختلفة لوسائط حمل المعلومات مثل النص والصوت والصورة(١)".

وتتكون بنية صفحات الويب من أجزاء أساسية يحتوي كل منها على معلومات ذات هدف مختلف عن الآخر، حيث تبدأ صفحة الويب أولاً برأس الصفحة page head الذي يتم التعبير عنه بـ<head> content </head> بما يحدد بداية ونهاية المعلومات. ويخصص رأس الصفحة لتحديد موضوع الصفحة ووصف الكلمات الدالة بها ، ويساعد هذا الجزء محركات البحث في التعرف إلى محتوى صفحة الويب. ثانيًا جسم صفحة الويب ويحتوي على المعلومات المراد بثها لمستخدمي الويب بما في ذلك من وسائط مختلفة لحمل

(١) ODLIS —Online Dictionary for Library and Information Science. OP.TC.

المعلومات. ويحدد هذا الجزء عنصرا الترميز <body> content </body>. ثم خاتمة صفحة الويب التي تحوي حقوق الملكية الفكرية وحقوق النشر.

إن ثمة جوانب لابد من مراعاتها عن بناء صفحات الويب صاغها اتحاد شركات الويب W3C ، وهي كالتالي[1]:

- أن يتسق تصميم صفحات الويب مع متطلبات مستخدمي صفحة الويب، حيث يراعي التصميم احتياجات المرحلة العمرية وجنس المستفيد والحالة المعيشية لمستخدمي الصفحات المتوقعين.
- سهولة الإبحار والتصفح داخل محتويات الصفحات والصفحات الداخلية المكونة لمواقع الويب.
- التأكيد على سلامة الروابط إلى الصفحات الأخرى داخل صفحة الويب، وعدم وجود روابط ضائعة Broken Links.
- التأكيد على جودة محتوى صفحات الويب بما يحقق كفاية أعلى داخل محركات البحث، فضلاً عن تحقيق أعلى درجة تصفح.
- أن يتم تكرار تحديث صفحات الويب بشكل دوري على مدد زمنية قصيرة.
- إمكانية التمتع بتقديم نشرات إخبارية E-News letters تذكر مستخدمي صفحات الويب من حين إلى آخر بزيارة هذه الصفحات.

---

(١) web designs. Realife web designs, 2005, cited 15/5/2007, cited at http://www.realifewebdesigns.com/web-resources/web-tips.html

٤ / ١ / ٢ / ٢ مواقع الويب:

عرف قاموس مصطلحات المكتبات والمعلومات على الخط المباشر ODLIS موقع الويب على أنه "مجموعة من صفحات الويب المتكاملة المتصلة معًا بروابط فائقة ومخزنة على أحد خادمات الويب، وتستخدم من خلال أحد برامج تصفح الويب. وهذه الصفحات متاحة ٢٤ ساعة في اليوم بحيث يمكن لمستخدمي الويب الاطلاع عليها في أي وقت. وصفحات الموقع هي نتيجة لعمل مؤسسة أو هيئة أو فرد على اختلاف اتجاهاتهم، ويطلق على شاشة الترحيب الأولى للموقع اسم الصفحة الدليلية Home page. ويجب على مصمم موقع الويب متابعة صفحاته لتحديثها من حين إلى آخر." ويتميز هذا المفهوم بالشمولية ؛ حيث تطرق إلى جوانب الإتاحة والتحديث والاسترجاع من خلال متصفحات الويب[١].

إن مواقع الويب وإن اختلفت في أهدافها وعدد صفحاتها ، فهي تسعى إلى اجتذاب أكبر عدد ممكن من مستخدمي الويب ، ولذلك فإن على مصممي مواقع الويب الاهتمام بمجموعة من خصائص مصادر المعلومات التي توفر للموقع أكبر عدد من الزيارات أعلى درجة أداء. مثل التي حددها Alastair Smith في حديثه عن تقييم مواقع الويب، وهي كما يلي[٢]:

---

(١) ODLIS —Online Dictionary for Library and Information Science. OP.TC.

(٢) Smith, Alastair G. Testing the Surf: Criteria for Evaluating Internet Information Resources." The Public-Access Computer Systems Review 8, no. 3 (1997). Cited 20/5/2007, cited at:
http://info.lib.uh.edu/pacsrev.html.

أولاً : مجال التغطية لموقع الويب: حيث يجب على الموقع توضيح حدود الموضوع داخله مثل:

- ❖ مدى الشمولية في عرض المعلومات.
- ❖ مدى العمق أو التخصص في الموضوع.
- ❖ تحديد المدة الزمنية التي تعالج فيها معلومات الموقع.
- ❖ طبيعة أشكال حمل المعلومات التي يحتوي عليها الموقع.

ثانيًا : محتوى موقع الويب كمصدر معلومات: حيث يجيب موقع الويب على أسئلةٍ كالتالية:

- هل يحمل الموقع معلومات أولية أم متخصصة في الموضوع؟
- هل يحمل الموقع المعلومات ذاتها أم روابط وإشارات ببليوجرافية عن الموضوع؟
- هل يتكامل الموقع مع مواقع أخرى متخصصة أم يقف بشكل فردي؟
- ما طبيعة المعلومات التي يحملها الموقع ، نصية أم صوتية أم مرئية؟

ويشمل الحديث عن محتوى موقع الويب الجوانب التالية: -

١- **دقة المعلومات:** هل يتبع هذا الموقع هيئة علمية متخصصة في الموضوع أم يعتمد الموقع على بقية مواقع الويب في استقاء المعلومات المعروضة داخله؟

٢- **المسئولية الفكرية**: فلمن تتبع هذه المعلومات؟

٣- **الحداثة :** وتتعلق بطبيعة موقع الويب في ثبات المعلومات أم إجراء التحديث الدوري لمحتوى صفحاته.

٤- **التفرد في المحتوى:** ويتعلق ذلك باختلاف الموقع عن غيره من مواقع الويب الأخرى في عرض معلومات الموضوع، أو أن يختلف موقع الويب في وسائط حمل المعلومات مع اتفاقه في المحتوى مع المواقع الأخرى.

٥- **الروابط مع المواقع الأخرى:** حيث يدعم الموقع الوصول إلى مواقع الموضوع الأخرى، فضلاً عن تحديث روابط صفحاته ذاتها وصيانتها من حين إلى آخر، أيضًا مدى حداثة مصادر المعلومات التي يرشد إليها هذا الموقع وتدل عليها روابطه إلى الصفحات الأخرى.

٦- **دقة التنسيق والصياغة:** حيث تهتم مواقع الويب بتنسيق المتن داخل صفحاتها، وفي حالة التحول إلى وسائط أخرى مثل المصادر المرئية فيجب على الموقع الاهتمام بمدى وضوح وكثافة المصادر المصورة.

ثالثًا : الرسومات والوسائط المتعددة :

يرتبط هذا العنصر بمدى قابلية المشاهدة لصفحات الموقع؛ إضافة إلى ملاءمة الأشكال المصورة لمحتوى الموقع وتوافقها مع موضوع موقع الويب.

رابعًا : الغاية والجمهور:

ويقصد بذلك أهمية التخطيط المسبق لبناء موقع الويب والتعريف بالهدف من تصميمه، أيضًا التوافق بين حاجات مستخدمي الويب وبين هدف ومقصد إنشاء موقع الويب.

خامسًا : تقييم المراجعات العلمية:

حيث يؤثر في تقييم مواقع الويب مدى الاستشهاد به وتقييمه في المراجعات العلمية، بالإضافة إلى تعليقات الهيئات العلمية والأفراد على مدى الإفادة من محتوى موقع الويب.

سادسًا : كفاية الأداء :

يهتم مستخدمو الويب بمدى الإفادة من استخدام مصادر الويب الرقمية، حيث إنه في المقابل سيترك المستخدم المصادر التقليدية هادفًا إلى القدر الأكبر من تحقيق كفاية استخدام مواقع الويب. وتقاس كفاية أداء الموقع لدى مستخدمي الويب من خلال مجموعة من العوامل هي:

١- قابلية الاستخدام وسهولته:

يهتم أغلب مستخدمي الويب بمدى سهولة تصفح مواقع الويب والبحث داخلها، إلى جانب ذلك مدى وضوح عناصر تشغيل الموقع والحصول على واجهة البحث الواضحة وأوامر التشغيل سهلة الاستيعاب.

٢- متطلبات التشغيل والتصفح:

يتساءل مستخدمو الويب دائمًا عن متطلبات تشغيل مواقع الويب من حيث احتياج هذه المواقع إلى برامج تشغيل خاصة أو متطلبات أخرى لشبكة الويب أو الاشتراك لدى الموقع أو الاحتياج لكلمات المرور، بالإضافة إلى نظم التشغيل المتوافقة مع موقع الويب أو خادم الويب.

٣- إمكانات بحث المعلومات:

ما إمكانات البحث التي يوفرها مصدر المعلومات؟ وما معاملات وآليات البحث داخل موقع الويب؟

٤- تنظيم وتصفح المعلومات:

يمكن لموقع الويب توفير تقسيمات موضوعية تسهل لمستخدمي الموقع التحرك بينه وبين فروعه، واسترجاع المعلومات المخصصة المرتبطة بمصطلحات البحث. فضلا عن اعتماد موقع الويب لتنظيم صفحاته بما يسهل للمستخدم تصفح محتوياته بسهولة تامة.

٥- مدى التفاعلية:

تتصف مواقع الويب بالتفاعلية عند تقديمها شاشات المساعدة ورسائل الحوار إلى مستخدمي الموقع عند الحاجة إليها. مما يمكن المستخدم من فهم آليات عمل موقع الويب والحصول على ما يريده من معلومات. على الجانب الآخر، فان التفاعلية تعني لبعض مستخدمي الويب مدى قدرة الموقع على تطوير وتحديث محتواه الموضوعي وقصر مدد التحديث، بما يكفل لمستخدمي الويب الحصول على آخر المعلومات في التخصص.

٦- الوصول والإتاحة:

يحتاج مستخدمو الويب إلى سهولة الاتصال بموقع الويب، وإمكانات تحميل محتوى الموقع بالسرعة المطلوبة، حيث تتصف بعض المواقع ببطء الوصول إليها وبطء تحميل محتواها على نوافذ الويب. هذا إلى جانب عنصر

الإتاحة المتمثل في الحصول على المعلومات بدون اشتراك أو دفع مقابل للمعلومات أو احتياج كلمات المرور.

سابعًا : التكلفة:

إن عامل التكلفة هو أهم عوامل الحكم على كفاية موقع الويب. حيث تقاس التكلفة من جانب مستخدمي الويب بعنصرين هما :

- ❖ تكلفة الاتصال بموقع الويب.
- ❖ تكلفة الحصول على المعلومات المتاحة داخل صفحات موقع الويب.

وتحتاج مواقع الويب إلى جانب معايير البناء والتصميم السابقة للمتابعة والرعاية الشاملة. وتتمثل جوانب رعاية مصادر الويب في العناصر التالية[1] :

١- متابعة التطوير والتحديث الدائم لموقع الويب بما يتماشى مع ما يجد من برامج تشغيل المصادر الرقمية، إلى جانب التوافق مع بيئات التشغيل المتطورة. فضلاً عن توفير عناصر أمن وحماية المعلومات ضد اختراقات مستخدمي الويب.

٢- متابعة الإجراءات الإدارية مع خادم الويب الذي يحتفظ بالموقع داخل قاعدة بياناته، وما يتبع ذلك من تحديث وتطوير لمسار موقع الويب domain name.

---

(١) Preservation Risk Management for Web Resources. D-Lib Magazine, 2002, cited 25/5/2007, cited at
http://www.dlib.org/dlib/january02/kenney/01kenney.html

٣- متابعة وصيانة موقع الويب على الشبكة بما يقي مصدر الويب التوقف نتيجة للضغط أو إساءة الاستخدام.

٤- متابعة سياسات النسخ والحفظ لمصدر المعلومات الرقمي داخل خادم الويب من حيث وسيط حمل النسخ الاحتياطية للموقع وأماكن تخزين هذه النسخ.

٥- الاهتمام بتوفير موقع جغرافي مناسب لخادم الويب الذي يحوي موقع الويب ومصادر المعلومات. وذلك بما يكفل لخادم الويب البعد عن الأخطار الطبيعية مثل درجات الحرارة والرطوبة والزلال.

٤ / ١ / ٢ / ٣ قواعد البيانات:

تتميز قواعد البيانات عن بقية مصادر الويب بخصوصية في التنظيم والبناء ومعالجة الملفات الرقمية داخلها، كما أن مصادر الويب تعتمد على العرض المباشر لما تحويه من معلومات خلاف قاعدة البيانات التي تتطلب استخدام واجهة ابحث ذات استراتيجيات البحث الخاصة لاستدعاء ملفاتها الرقمية. وقد عرف قاموس مصطلحات المكتبات والمعلومات ODLIS قاعدة البيانات على أنها " مجموعة كبيرة من الملفات تحوي معلومات رقمية في أشكال مختلفة مثل البيانات البيليوجرافية والمستخلصات والصور والنصوص الكاملة... وغيرها. ينتمي جميعها إلى مجال موضوعي محدد وترتكز في بنيتها على مجموعة التسجيلات المعيارية بحسب طبيعة قاعدة البيانات، وذلك حتى يتم تخزينها وتنظيمها واستدعائها من خلال نظم إدارة قواعد البيانات DBMS. وتتنوع قواعد البيانات داخل مجال المكتبات بين قواعد البيانات الببليوجرافية

كالفهارس، وقواعد بيانات المستخلصات ، وقواعد بينات النصوص الكاملة[1]."

ويكشف المفهوم السابق عن أهمية برامج إدارة قواعد البيانات ومدى تأثيرها في عزل قواعد بيانات الويب غير المرئية عن محركات البحث، ما يخفيها بالتالي عن أعين مستخدمي الويب. ويبرز هنا مدى الحاجة إلى التوافق بين آليات عمل نظم إدارة قواعد البيانات وبين آليات عمل محركات البحث حتى يتمكن محرك البحث في مرحلة لاحقة من التعمق داخل محتوى قواعد البيانات واستدعاء ملفاتها ضمن نتائجه أمام مستخدمي الويب. ومن بين أشهر أنواع نظم إدارة قواعد البيانات نظامًا Microsoft's Access and Borland's Paradox وهما شائعا الاستخدام في أجهزة الحاسب الشخصية PCs، أيضًا فهناك نظم إدارة قواعد البيانات ذات الحجم والتعقيد العالي مثل أنظمة Oracle 8 or Sybase SQL or SAP.

ويساهم العنصر البشري أيضًا في محاولة رفع كفاية الأداء لقاعدة البيانات فيما يطلق عليه Database Administrators، وتحدد وظائف مديري قواعد البيانات على النحو التالي[2]:

١- صيانة وتحديث نظم إدارة قاعدة البيانات.

٢- بناء وتقييم سياسات عمل قاعدة البيانات.

٣- توظيف وتدريب القائمين على عمل قاعدة البيانات.

---

(١) ODLIS —Online Dictionary for Library and Information Science. OP.TC.

(٢) Database Administrator Jobs and information. UniXL , 2004, cited 2/6/2007, cited at http://www.unixl.com/dir/information_technology/databases/

٤- الاستعداد التام لتغيير نظم العمل وتطوير التكنولوجيا المستخدمة.

٥- متابعة تقارير عمل قاعدة البيانات.

٦- تحليل كفاية عمل قاعدة البيانات وجودة المعلومات المستخرجة.

٧- تحديد متطلبات المستفيدين الجدد وتنمية كفاية قاعدة البيانات.

لقد عملت قواعد البيانات منذ بدايتها على توفير ميزات افتقدتها أشكال التخزين التقليدية؛ مثل توفير سرعة الحصول على المعلومات، وتخزين أكبر قدر من البيانات، إضافة إلى عمليات التحديث والبحث عن المعلومات.وعلى ذلك يمكن صياغة مهام قاعدة البيانات في الحفظ والتنظيم بجانب سرعة البحث ودرجة الكفاية. وجاءت الويب لتحمل محتوى مصادر قواعد البيانات إلى مستخدمي الويب بواسطة الاتصال بالإنترنت. ومن ثم تعد قاعدة البيانات أحد مصادر المعلومات الرقمية المنتمية للويب. عملت قواعد البيانات كمصدر للمعلومات الرقمية على إحلال نظام تكشيف الربط اللاحق بدلاً من تكشيف الربط المسبق حتى يتمكن المستخدم من استرجاع تسجيلات قاعدة البيانات بواسطة كلمات مفتاحية داخل مختلف حقول التسجيلة. غير أن هناك من رأى أن لهذا الأسلوب في التكشيف بعض العيوب مثل[١]:

- ارتفاع نسبة الاستدعاء، حيث يتم استرجاع تسجيلات كثيرة مشتركة فقط في بعض الكلمات الدالة.

---

(١) What is a Database. :Introduction to Databases for Web Developers Extropia, 2007, cited 25/5/2007, cited at http://www.extropia.com/tutorials/sql/what_is_database.html

- تحتاج بعض نظم إدارة قواعد البيانات إلى درجة التحكم العالي ودقة تغيير وبحث البيانات.
- ارتفاع تكلفة بعض برنامج تشغيل قواعد البيانات.

غير أن الباحث يرى أن مختلف هذه العيوب لم تعد بالشكل المذكور سابقًا نظرًا لما شهدته نظم إدارة قواعد البيانات من تطوير.على الجانب الآخر فإن قواعد بيانات الويب تقسم إلى نوعين، وهما[1]:

أولاً : قواعد البيانات التحليلية: OLAP- On Line Analytical Processing

وهي قواعد البيانات التي يتم إنشاؤها لاختزان معلومات محددة والاستعانة بها بحسب الحاجة إليها ، غير أن الدور الرئيس لقاعدة البيانات يتمثل في عمليات التحليل وإنتاج معلومات جديدة تدل على طبيعة واتجاهات البيانات السابق تخزينها. ومن أمثلة قواعد البيانات التحليلية على الويب قاعدة بيانات Amazon.com ، حيث يعمل برنامج إدارة قاعدة البيانات على تحليل أسئلة بحث مستخدمي الويب ثم إنتاج الواجهة المحددة لنتائج البحث.

ثانيًا : قواعد معالجة العمليات: OLTP On Line Transaction Processing)

حيث يقوم هذا النوع بمهام قاعدة البيانات من تخزين واسترجاع للبيانات، إلى جانب تقارير حالة المعلومات من حذف وتعديل وإضافة.

بجانب النوعين السابقين، فان قواعد البيانات تعتمد في بنائها على ثلاثة نماذج، يعتمد أحدها بحسب طبيعة عمل قاعدة البيانات والهيئة المنشئة، وهي :

(١) Introduction to Databases for Web Developers. Lbid.

٤ / ٢ المعالجة والتنظيم لمصادر الويب :

يشمل الحديث عن المعالجة الفنية والتنظيم لمعلومات الويب شقين أساسيين هما ؛ المعالجة الفنية للمعلومات في بيئة الويب ، والمعالجة والتنظيم لمصادر الويب الرقمية. وهي كالتالي:

٤ / ٢ / ١ المعالجة والتنظيم لبيئة الويب :

افتقدت المعلومات الرقمية على الويب إلى شكل المعالجة والتنظيم الكامل داخل مصادر الويب، وقد حظيت هذه المشكلة بالدراسة من باحثي علوم الحاسبات والمعلومات. واتفق الباحثون على وضع تدرج منطقي لأسباب هذه الظاهرة يتضح في العرض التالي(١):

- إن المعلومات الرقمية على الويب دائمة التغيير بتغير الوقت.
- تحتاج المعلومات المتشابهة في الموضوع مختلفة المصادر إلى الترابط فيما بينها.
- يحتاج ربط المعلومات المتشابهة معًا إلى نظام تحليل دقيق لعملية تنظيم هذه المعلومات.
- يحتاج نظام التحليل والتنظيم الدقيق إلى رسم واضح لمهامه وإجراءاته في تنظيم المعلومات على الويب.
- إن إجراءات ومهام تنظيم المعلومات دائمًا ما تتصف بالثبات في الوقت الذي تحتاج فيه إلى التطوير الدائم.

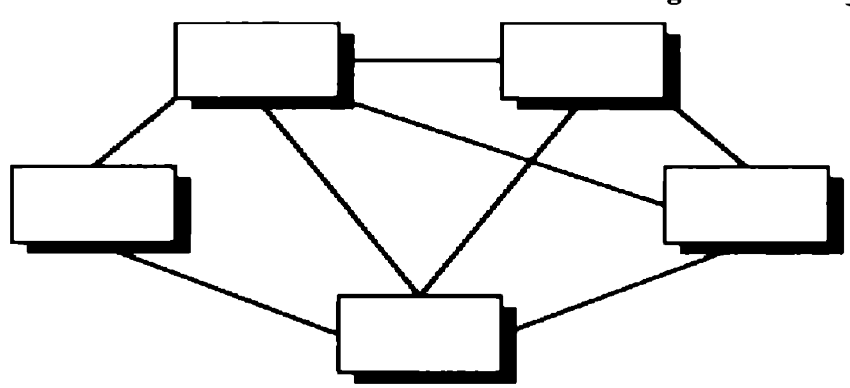
---

(١) Ke-Thia Yao, Robert Neches. Op.tc.

وتتضمن الويب نماذج عامة ذات أشكال متعددة لحمل ولتنظيم المعلومات، وهذه الأشكال يمكن التعرف إليها كالتالي[1]:

٤ / ٢ / ١ / ١ نموذج الويب The Web Model:

تبنى المعلومات داخل هذا النموذج باستخدام كلمات ذات دلالة في كل صفحة ويتم ربط الصفحات فيما بينها عن طريق روابط بين هذه الكلمات، والانتقال من صفحة إلى أخرى بالضغط بشكل عشوائي يأتي من متابعة كلمات محددة ولا يرتبط بتتابع الصفحات أو التحرك داخل مستوى واحد لبناء موقع الويب. تستخدم الويب لغة خاصة تسمى **HTML** تأتي من مصدر عام للغات ترميز النصوص الفائقة هي لغة الترميز المعيارية العامة **SGML**. وتتكون الوثيقة الرقمية في هذه الحالة من عناصر يعبر عنها بحقول أو محارف الصفحة **HTML Tags**.



الشكل رقم (٤ / ٢) نموذج الويب لتنظيم المعلومات

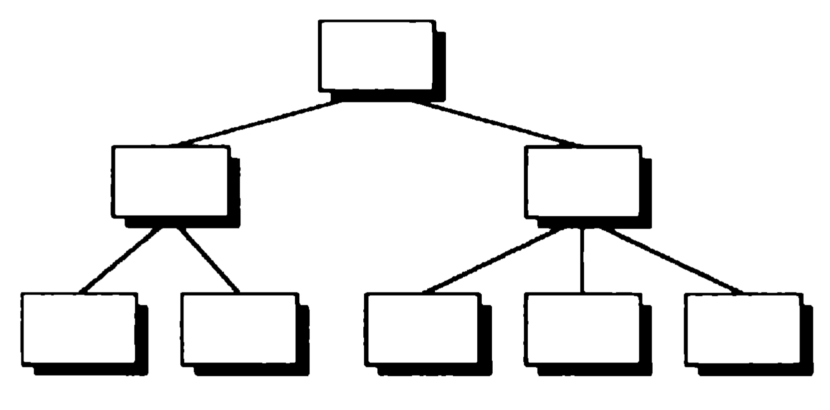
---

(١) Watson, James. Information Models for Web Structure, published by NIVA Inc. cited 20/5/2007, Cited at http://www.writersblock.ca/spring1998/feature.htm.

ويعتمد مستخدمو الويب في ذلك على أسلوب التصفح واستخدام الروابط الفائقة في استرجاع المعلومات داخل هذا النموذج. غير أن ذلك يتطلب من مستخدمي الويب مهارات تذكر الصفحات وكفاية التحرك بينها إلى جانب الانتقال اليسير بين مستويات عرض المعلومات داخل صفحات الموقع الواحد للتعرف إلى مختلف المعلومات المتاحة. وأخيرًا فإن بنية الروابط الفائقة وعدم التزامها بمسارات شجرية أو متوازية يجعل من الصعب صيانتها وتحديثها. ويتمثل هذا النموذج في شكل صفحات الويب ومواقعها على اختلاف تخصصاتها أو وسائط حمل المعلومات بها.

٤ / ٢ / ١ / ٢ النموذج الشجري أو الهرمي: The Tree Model

يتمثل النموذج الشجري لتنظيم المعلومات في بناء دليل يضع المعلومات المتعلقة بالموضوع أو موقع الويب في مستويات ترتب من العام إلى المتخصص فالأكثر تخصصًا. وهو ما يسهل على مستخدمي الويب تحديد موضعهم داخل موقع الويب في أثناء بحث واستدعاء المعلومات.

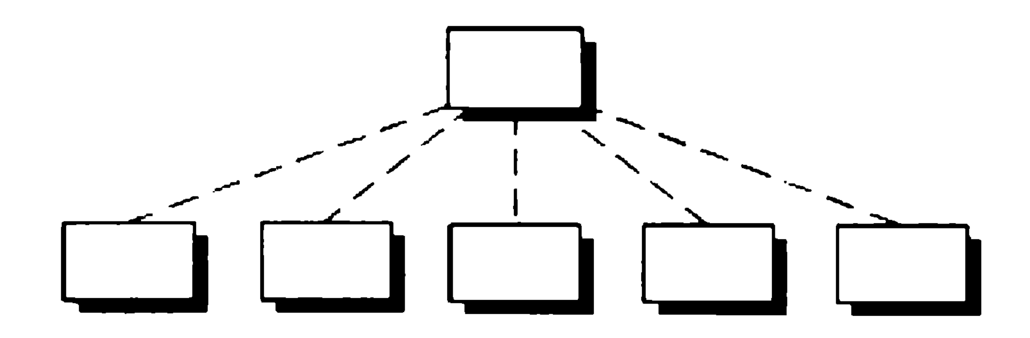


الشكل رقم (٤ / ٣) النموذج الشجري لتنظيم المعلومات

تعتمد مواقع الويب في هذا النموذج على طبيعة الموضوع ذاته وتفريعاته في تنظيم معلومات الويب؛ حيث تتميز بعض قطاعات المعرفة بوجود قوائم لرؤوس الموضوعات المحددة لطبيعة المجال أو الموضوع تحصر مختلف موضوعاته الفرعية. غير أن هذا النموذج يحتاج نوعًا من مستخدمي الويب يتميز بالتخصص الموضوعي إلى جانب قدرته على التحرك بين تفريعات الموضوع العامة والخاصة. ويمثل هذا النموذج أدلة بحث الويب search directories التي تقسم المعرفة أو قطاعات منها إلى فروع أخص يمكن لمستخدمي الويب التحرك بين فروع المعرفة أو القطاع المعرفي الواحد.

٤ / ٢ / ١ / ٣ نموذج بحث المعلومات The Searchable Model

أفرز التقدم في أجهزة وبرامج الويب ما يسمى بنموذج بحث المعلومات. وفي هذا النموذج لا يحتاج مستخدم الويب التنقل بين قطاعات أو فروع الموضوع، وإنما الاعتماد على استدعاء المعلومات المطلوبة مباشرة . إن مصطلح محركات البحث هو المسمى الآخر لهذا النموذج من نماذج تنظيم معلومات الويب؛ حيث يحوي محرك البحث قاعدة تضم نسخًا وتسجيلات عن صفحات ومواقع الويب يتم استرجاعها باستخدام الكلمات المفتاحية. ومن أمثلة ذلك **Google** و **AltaVista**.



الشكل رقم (٤ / ٤) نموذج بحث المعلومات لتنظيم الويب

يتصف هذا النموذج بمجموعة من السلبيات التي تعوق تحقيق الفائدة الكاملة ومنها:

١- إن هذا النموذج يهتم غالبًا بتحقيق نسبة الاستدعاء أكثر من الاهتمام بنسب التحقيق للنتائج المسترجعة، مما سبب ضياع وقت وجهد مستخدمي الويب في استرجاع المعلومات.

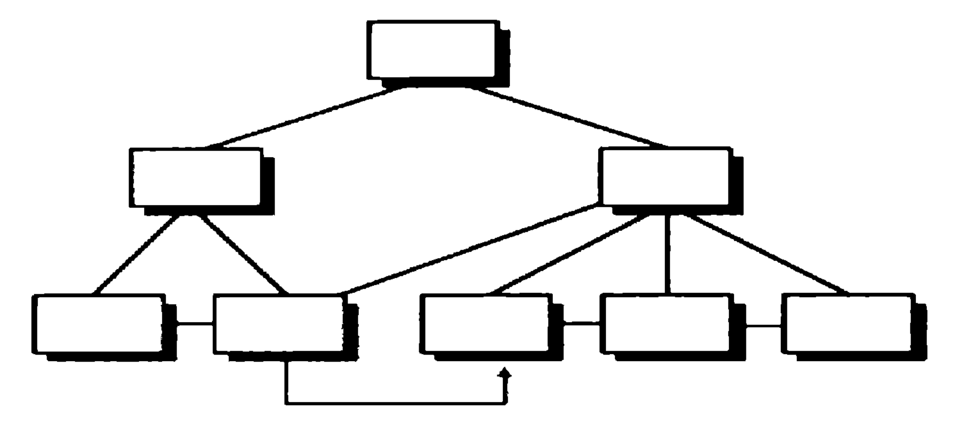
٢- يستطيع هذا النموذج إجادة التعامل فقط مع ملفات المعلومات المصاغة بلغة html دون غيرها من أنواع الملفات الأخرى، مما يتسبب معه في ضياع كم هائل من المعلومات في نطاق اهتمام مستخدمي النموذج.

٣- يعتمد هذا النموذج في ترجيح المعلومات المسترجعة على حساب تكرار وتردد بعض الكلمات الواردة في صفحات ومواقع الويب، غير أن ذلك يتسبب في أحيان كثيرة في استرجاع صفحات ومواقع لا ترتبط بموضوع البحث.

٤- يحتاج نموذج محركات البحث إلى ضبط الدلالة بين كلمات العناوين في صفحات وموقع الويب وبين ما تدل عليه هذه الكلمات من موضوعات.

٤ / ٢ / ١ / ٤ نماذج المعلومات المتداخلة أو المهجنة : The Hybrid Models

تحاول بعض مواقع الويب الجمع بين أشكال النماذج السابقة في نظام متداخل أو مهجن؛ حيث تستخدم تقسيمًا موضوعيًا يساعد مستخدمي الويب على التعرف إلى جانب الموضوعات العامة والخاصة، أيضًا استخدام الوصول السريع ببحث الكلمات الدالة، بالإضافة إلى إمكانية تصفح الموقع باستخدام الروابط الفائقة.



الشكل رقم (٤ / ٥) نموذج بحث المعلومات المتداخل أو المهجن

إن استخدام هذا النموذج يتطلب من مصممي مواقع الويب دقة صياغة وتكامل أساليب تنظيم وبحث المعلومات فيما بينها ، إلى جانب توفير نقاط المساعدة لمستخدمي الويب في الانتقال من أسلوب إلى آخر. وأفضل الأمثلة على هذا النموذج مواقع بوابات الويب (web portals (gateways.

٤ / ٢ / ٢ المعالجة والتنظيم لمصادر معلومات الويب :

٤ / ٢ / ٢ / ١ معالجة وتنظيم صفحات الويب :

ترتقي صفحة الويب إلى رتبة الموقع في حين تقف بمفردها كمصدر مستقل حمل بيانات الويب على اختلاف أشكالها ؛ حيث تصبح في هذه الحالة موقعًا على الويب من صفحة واحدة. وتتلقى صفحات الويب المعالجة الفنية وبنية التنظيم منذ عملية إنشاء وتصميم صفحة الويب، وسوف يقصر الحديث هنا على معالجة صفحات الويب في شكلها الطبيعي الذي تكون فيه وحدة متكاملة مع صفحات أخرى لتكوين موقع الويب.

إن ثمة طريقتين يمكن بهما معالجة وتنظيم صفحات الويب وهما:

٤ / ٢ / ٢ / ١ / ١ المعالجة الفنية :

تستخدم الويب عناصر الميتاداتا metadata لوصف ملفاتها ومصادرها الرقمية عامة ومعالجة صفحات ومواقع الويب خاصة. وقد لاحظ الباحث لما شاهده من استخدام الويب أن المعالجة الفنية لملفات الويب تأتي بتخصص أدوات التنظيم والبحث ؛ حيث تهتم أدوات بحث الويب من محركات وأدلة بواحد أو أكثر من ملفات الويب محددة الشكل، فتوجد محركات بحث الويب ومحركات الوسائط المتعددة على إفراد أشكالها. وعلى ذلك فان المعالجة الفنية لصفحات الويب تدخل ضمن نطاق أدوات بحث الويب. إلا أن صفحات ومواقع الويب لم تلق الاهتمام الكافي بصياغة معايير المعالجة الفنية داخل أدوات البحث كما حظيت بها الملفات الرقمية على اختلاف وسائط حمل المعلومات.

تبدأ معالجة صفحات الويب مع بداية كتابة وتصميم محارف بناء صفحة الويب html tags ؛ حيث تصاغ بعض حقول وصف الصفحة ومحتواها بعناصر مثل meta key word , meta title بما يعكس معه محتوى صفحة الويب. غير أن هذه العناصر دائمًا ما تحمل أهدافًا واحتياجات مصمميها لتحقيق أكبر عدد من زيارات مستخدمي الويب للصفحة، بما يدفعهم إلى وضع كلمات ذات دلالة على موضوعات أخرى أكثر جاذبية لمستخدم الويب وإن بعدت أو حتى اختلفت تمامًا عن موضوع صفحة الويب الأساس. يأتي بعد ذلك دور محركات وأدلة البحث في معالجة صفحة الويب ؛ حيث تعتمد محركات البحث بشكل

أكبر من الأدلة على استخدام محارف html tags في تنظيم صفحات الويب داخل قاعدة بياناتها. أيضًا يؤدي متن صفحة الويب دورًا مهمًا في اقتباس عناصر معالجة صفحة الويب داخل محركات البحث ، ذلك لاعتماد محركات البحث على آلية التنظيم والمعالجة بموقع الكلمات ومدى ترددها أو تواترها بكثرة داخل صفحة الويب. على الجانب الآخر ، فإن أدلة البحث تنتهج استخدام العنصر البشري في رسم الصورة الحقيقية للتعبير عن المحتوى الموضوعي لصفحة الويب ؛ حيث تصاغ الكلمات الدالة المعبرة عن الموضوع الحقيقي لصفحة الويب، إلى جانب ربط صفحة الويب بقائمة رؤوس الموضوعات داخل دليل البحث؛ ليتم استدعاؤها عن طريق تصفح تفريعات الموضوع داخل دليل البحث.

لقد أكدت بعض الدراسات مثل Edward T O'Neill أن استخدام عناصر الميتاداتا في ازدياد مستمر داخل صفحات ومواقع الويب. وعملت هذه الدراسة على قياس نسبة استخدام حقول الميتاداتا داخل مواقع وصفحات الويب في الفترة من ١٩٩٨- ٢٠٠٢م وقد عكست نتائج الدراسة أن أهمية واستخدام حقول المعالجة الفنية دائمًا في تزايد ، وأوضحت النتائج سعي صفحات ومواقع الويب إلى الاستعانة بمختلف معايير الوصف على الويب مثل Dublin core لوصف ملفات الويب وذلك كما في الجدول رقم (٤ / ٢) (١):

---

(١) O'Neill, Edward T. Trends in the Evolution of the Public Web, D-Lib Magazine, April 2003, cited 20/6/2007, cited at http://www.dlib.org/dlib/april03/lavoie/04lavoie.html.

الجدول رقم (٤ / ٢) نسب استخدام الميتاداتا داخل صفحات ومواقع الويب

Table 1: Metadata Usage on Public Web Sites, 1998 – 2002					
	1998	1999	2000	2001	2002
عدد المواقع العامة المستخدمة للميتاداتا	1,457,000	2,229,000	2,942,000	3,119,000	3,080,000
نسبة المواقع المستخدمة للميتاداتا في صفحاتها الدليلية home page	70 percent	79 percent	82 percent	85 percent	87 percent
Mean No. of Tags عدد حقول الوصف	2.75	2.32	2.72	2.97	3.14
% of pages with Metadata	45 percent	50 percent	59 percent	62 percent	70 percent
Mean No. of Tags	2.27	2.16	2.46	2.63	2.75

٤ / ٢ / ٢ / ١ / ٢ التنظيم :

إن الطريقة المثلى في تنظيم صفحات الويب هي الربط الفائق hyper linking، خاصة عندما تكون صفحة الويب جزءاً من موقع يضم مجموعة كبيرة من صفحات الويب. ومن ثم يكون التنظيم العشوائي واستخدام النص الفائق هو أحد أسلوبين لتنظيم صفحات الويب. يأتي الأسلوب الثاني متمثلا في عمل تسجيلات قواعد بيانات محركات البحث أو أدلة البحث على الويب، حيث تنظم صفحات الويب بحسب محارف html الواصفة للمحتوى والصفحة. وتكون هذه التسجيلات أكثر إحكامًا وتنظيمًا في أدلة البحث عنها في

محركات البحث ؛ حيث الوصف البشري المعتمد على معيار ثابت لوصف الويب حتى وإن كان معيارًا محلي الاستخدام داخل دليل البحث in home use.

٤ / ٢ / ٢ / ٢ معالجة وتنظيم مواقع الويب:

٤ / ٢ / ٢ / ٢ / ١ المعالجة الفنية:

تعاني مواقع الويب من افتقار معايير المعالجة والتنظيم، سواء أكان ذلك في بداية التصميم أم في داخل محركات البحث، حيث ترصد شركة Website-Designer.net أنه من بين مليار موقع على الويب تمكنت أدوات بحث الويب من الوصول إلى ٥٪ فقط ، إلى جانب ٥٪ تم تكشيفها باستخدام معايير معالجة فقيرة بعض الشيء. وقد رأت الشركة أيضًا بعض جوانب الاتفاق في عناصر المعالجة بين المطبوعات ومواقع الويب مثل ؛ اختيار المادة المكشفة ، الوصف الشكلي للمصدر، طبيعة التكشيف في المنفردات أو السلاسل، طبيعة المداخل الموضوعية[(١)].

يرى الباحث أن هناك تداخلاً في واقع التطبيق بين مصطلحات التكشيف والوصف الفني والمعالجة والتنظيم في البيئة الرقمية عنها في البيئة الورقية التقليدية ؛ حيث يختص كل مصطلح من هذه المصطلحات بأداء عملية مختلفة تمامًا عن الأخرى في البيئة التقليدية، غير أنه لا يمكن الحديث عن فارق في جوانب الوصف المادي والموضوعي في المصادر الرقمية عندما تخضع كامل حقول التسجيلة الرقمية للبحث والاستدعاء بواسطة الكلمات المفتاحية. مثالاً

---

(١) Website Indexing. Website-Designer, 2007, cited at 6/6/2007, cited at http://www.website-designer.net/main.php.

على ذلك التكامل في استراتيجية البحث بين كلمتي " الحجم k.b١٦ و المياه" عند استرجاع ملف صورة رقمية أحد خصائصه المادية الوزن ١٦ k.b والمحتوى الموضوعي يتعلق بالمياه. وعلى ذلك فإن المعالجة الفنية والتنظيم لمواقع الويب وصفحاتها تعتمد كلية على استخدام مجموعة من الكلمات الواصفة مكونة حقول التسجيلة الرقمية التي تخضع لآلية بحث الكلمات المفتاحية. ويعكس حقيقة ذلك اعتماد مواقع الويب ومصمميها على محركات البحث بشكل تام في وضع المعالجة الفنية لمواقع الويب داخل قواعد بياناتها ، ومن ثم تعد محركات البحث هي أدوات المعالجة والتنظيم لمواقع وصفحات الويب.

لقد أجرت جامعة University of North Carolina دراسة على عينة عشوائية لمواقع الإنترنت بلغت ٢٧٠٠ موقع تم اختيار ٤٦٤ منها لدراسة ملامح التسجيلات الفنية بها ، وإلى جانب تأكيد الدراسة على افتقار مواقع الويب عامة لمعايير المعالجة الفنية ، فقد عملت الدراسة على تحديد الأطر العامة التي تتسم بها تسجيلات الوصف الحالية ، وخرجت نتائج الدراسة بأحد عشر ملمحًا أدرجت أسفل ثلاثة قطاعات لوصف مواقع الويب وصفًا فنيًا. يوضحها الجدول رقم (٤ / ٣). بالإضافة لذلك نبهت الدراسة القائمين على معالجة المواقع بأهمية الواقع التفاعلي والديناميكي لمواقع الويب ، حيث يتسم محتوى مواقع الويب بالتغيير والتحديث الدائم ، مما ينعكس على تنظيم حقول التسجيلات الرقمية[1]:

---

(١) Luo, Lili and West, David and Marchionini, Gary. A Study of Annotations for a Consumer Health Portal , 2006, cited 20/6/2007, cited at http://ils.unc.edu/annotation/publication/luo_jcdl_2005.pdf.

الجدول رقم (٤ - ٣) ملامح وجوانب الوصف لمواقع الويب

Facet	Category
Content المحتوى الموضوعي	Website Navigation: الموضوعات المتعلقة بتصفح موقع الويب
	Categorization of Geographic Scope : الجوانب المتعلقة بالرسومات التوضيحية داخل المواقع
	Categorization of Topical Scope: مناقشة العناصر الموضوعية التي يركز عليها موقع الويب
	Miscellaneous: مجموعة العناصر التي تتعلق بالموضوعات المختلفة داخل موقع الويب
Format الشكل	Question: Any comments that are questions along with anything that can reasonably be inferred to be a question.
	Answer: Any comment explicitly in response to a question and comments which probably answer unasked questions
	Statement: معلومات حول الموقع
Function أداء موقع الويب	Log of Action: تقرير عن حدث أو إجراء تم في الماضي
	Reminder : مجموعة من العناصر التي تحدد سير العمل في موقع الويب اعتمادا على ما تم وما سيكون لعمل موقع الويب فيما بعد
	Reach Consensus: A statement made in the process of reaching an agreement on a disputed point.
	Action Request: وهو الوقت الذي يأخذه موقع الويب لتحقيق التفاعل بينه وبين المستخدم في تنفيذ المهام المختلفة

يوضح الجدول رقم (٤ / ٣) القطاعات الأساسية الثلاثة لبناء التسجيلات الرقمية لمواقع الويب، وقد أدرجت أسفل كل منها العناصر المتعلقة بكل جانب من جوانب الوصف. وتعكس هذه العناصر الملامح التالية:

أولاً : بنيت التسجيلة الرقمية على عناصر ثلاثة هي المحتوى content والشكل format والمهام أو الأحداث Function . وقد عكس ترتيب هذه

العناصر في الجدول الاهتمام الكبير بمحتوى موقع الويب، والرسالة المعلوماتية المطروحة داخل محتوى موقع الويب الذي يعتبر وسيلة التنظيم والاسترجاع الأساسية لمواقع الويب. أعقب ذلك الحديث عن الشكل ثم المهام أو الأحداث الفنية والتقنية لموقع الويب.

ثانيًا : كان أول عناصر التسجيلة هو المحتوى المعبر عن الاتجاهات الموضوعية المختلفة لمواقع الويب، وظهر ذلك في عناصر :

١. Website Navigation : ويتضمن التقسيم الموضوعي الذي بني عليه موقع الويب.

٢. Categorization of Geographic Scope : الجانب المكاني أو الجغرافي الذي يركز عليه محتوى موقع الويب.

٣. Categorization of Topical Scope : ويغطي الموضوع أو القطاع المعرفي الآتي في متن أو وسائط موقع الويب.

٤. Miscellaneous : يغطي هذا الحقل الجوانب الموضوعية التي لم تذكر في العناصر السابقة ويرى واصف موقع الويب أهميتها في التعبير عن محتوى الموقع.

ثالثًا : لم يحتو عنصر الشكل format على خصائص الوصف المادية، كما نراه في تسجيلات الوصف الفني عامة، غير أن ذلك يتفق واقعيًا مع موقع الويب كمصدر معلومات لا توجد خصائص أو ملامح مادية. واقتصر عنصر الشكل هنا على ذكر أهم التعليقات والحقائق عن الموقع التي يمكن أن تطرح كأسئلة بحث تستخدم في استرجاع موقع الويب.

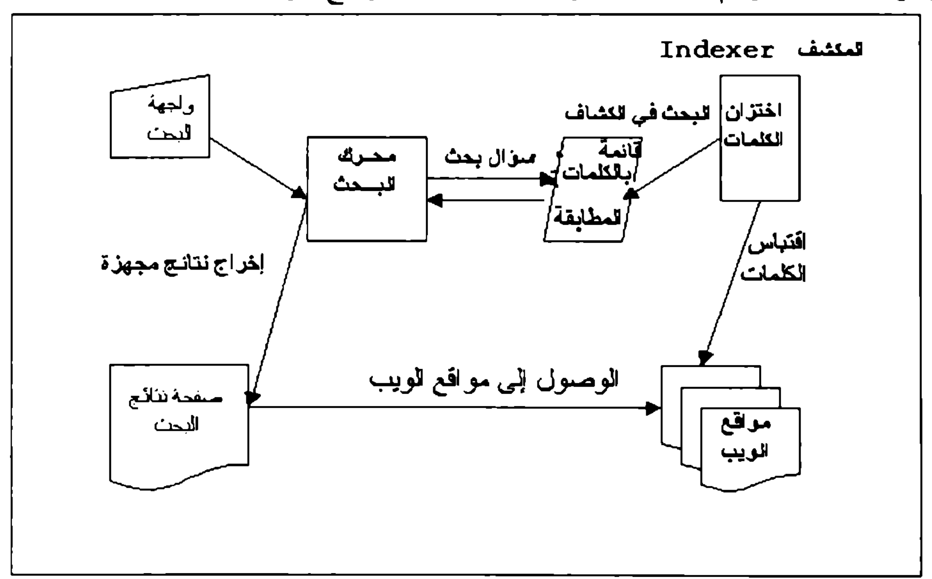
رابعًا : اهتم العنصر الثالث المهام Function بتاريخ ومستقبل موقع الويب من الأحداث أو التغيير التي مر أو سيمر بها موقع الويب. يعكس ذلك عنصر Log of Action المخصص لسرد تفاصيل التغييرات التي مر بها الموقع، وعنصر Reminder لتحديد التغييرات التي لابد من إحداثها للموقع في المستقبل، وعنصر Reach Consensus الذي يقترب من الحديث عن المسؤولية والملكية الفكرية لموقع الويب، وختامًا كان عنصر Action Request الذي يعتبر تعليقًا من القائم على معالجة ووصف موقع الويب، يذكر فيه أهم التعليقات والخصائص التي يراها في الموقع.

٤ / ٢ / ٢ / ٢ / ٢ تنظيم مواقع الويب :

تعد محركات وأدلة بحث الويب وحدها الأداة المنوطة بتنظيم مواقع الويب ومختلف الملفات الرقمية الأخرى. وعندما نتناول تنظيم مواقع الويب؛ فإننا نتحدث بالفعل عن عملية التكشيف والتحليل داخل محركات وأدلة البحث ؛ حيث تستخدم محركات البحث أحد برامجها " المكشف Indexer" لتحليل محتوى ونص مواقع وصفحات الويب باستخدام آليات مختلفة لاستخراج الكلمات الدالة المعبرة عن محتوى موقع الويب ؛ فتعتمد محركات البحث مثلاً عدد تردد كلمات محددة في نص صفحة الويب أو ظهور بعض الكلمات في أماكن التعريف بالمحتوى داخل صفحة الويب مثل محارف meta key word, meta content. وعلى ذلك تقف الدراسات عند تناول عملية التنظيم والتحليل داخل محركات وأدلة البحث دون الحديث عن تنظيم مواقع الويب ذاتها ، ذلك أن بداية ونهاية عملية التنظيم تتم داخل قاعدة بيانات أدوات البحث باستخدام البرامج أو العنصر البشري.

تعد عملية تنظيم وتحليل محتوى مواقع الويب واسطة العقد بين عمليات ثلاث تتم داخل أدوات بحث الويب. وتتم عملية التنظيم لمحتوى مواقع الويب من خلال تكشيف متن صفحات الموقع والخروج بأهم الكلمات الدالة على موضوع أو موضوعات موقع الويب. فبعد أن تنتهي مرحلة إضافة نسخة من موقع الويب إلى قاعدة بيانات محرك البحث، يبدأ عمل البرنامج الثاني في أدوات البحث الذي يعرف بالمكشف أو المفهرس، حيث يقوم بإعداد تسجيلات ببليوجرافية لمواقع الويب حديثة الإضافة إلى قاعدة بيانات أداة البحث، وتعمل هذه التسجيلة وفقًا لما ورد في المكتبات والمعلومات عن (التمثيل المكثف) حيث لا يتم اختزان موقع الويب بأكمله في عدد من أدوات البحث، وإنما حقول وبيانات منها مسار هذا الموقع للربط به والوصول إليه، وكل ذلك في شكل تسجيلة كما هو الحال في الفهارس الآلية للمكتبات OPAC. وتتم عملية الفهرسة في حالة الأدلة بشكل يدوي، أما في حالة محركات البحث، فإنها تتم آليًا، وفي حالة التكشيف والفهرسة الآلية؛ فإن برامج التكشيف الآلي تستخدم بنية صفحة الويب في التعرف إليها ووضع كلمات أو مصطلحات معبرة عن الموضوع أو الموضوعات التي تضمها صفحة الويب؛ حيث يفحص البرنامج هنا كل أجزاء الصفحة التي تحتوي على كلمات تخص الموضوع من كلمات العنوان، وكلمات النص، والكلمات المفتاحية التي وضعها مصمم صفحة الويب في رموز meta tags وتجمع هذه الرموز كل الكلمات التي يرى مصمم الصفحة أنها تعبر عن موضوع أو موضوعات صفحة الويب. وبعد تفحص كل الجوانب السابقة يتم وضع كلمات مفتاحية داخل تسجيلة هذا الموقع أو وضع كل كلمات النص

الطبيعية في وضع نشط؛ ليتم مضاهاتها بعد ذلك مع كلمات البحث التي يضعها الباحث عند إجراء البحث. وتختلف مواضع التكشيف من برنامج إلى آخر؛ حيث تعمل بعض برامج التكشيف على الاكتفاء بالجزء الأول فقط من بنية اللغة المعيارية لصفحة الويب، كما تعمل بعض البرامج على تكشيف عناوين الصفحات بجانب مسار الصفحة، وتمتد قدرة برامج التكشيف في أحيان أخرى إلى تكشيف كامل متن صفحة الويب وهو ما يجعل محركات البحث في هذه الحالة تمتاز بوفرة في الاستدعاء مع انخفاض نسبة التحقيق. ويبين الشكل رقم (٤ -٦) مرحلة تكشيف مواقع الويب[١]:



الشكل رقم (٤ -٦) مرحلة تكشيف مواقع الويب

(١) سيد ربيع سيد. **محركات بحث الصور على الإنترنت : دراسة تحليلية لوضع مواصفات محرك بحث صور.**؛ إشراف هانئ محيى الدين عطية ، جامعة بني سويف، ٢٠٠٥، أطروحة ماجستير. ص ٢٠.

وقد أشارت Laura Cohen إلى بعض الجوانب التي تؤخذ في الاعتبار عند تنظيم مواقع الويب، مثل[1]:

- تتصف بعض مواقع الويب بسرعة التغير في المحتوى نظرًا للتحديث السريع، ولذلك تتغير الكلمات الدالة على أثر تغيير المحتوى.
- إن إمكانات النشر التي توفرها حزمة برامج المكتب office تعطي مواقع الويب صلاحية النشر غير المقنن مختلف الأشكال، مما يؤثر بدوره على تقنين عملية معالجة وتنظيم مواقع الويب.
- إن الوصول إلى سياسات النشر لدى مواقع الويب غالبًا ما يعتمد على المصادفة ؛ حيث تحرص مواقع الويب على الاحتفاظ بسياسة نشر وتصميم مواقع الويب حتى تحتفظ بتميزها أو خصائصها عن المواقع الأخرى.

وقد أضافت الكاتبة أيضًا بعض المراحل المقترحة لتنظيم مواقع الويب، يمكن استخدامها كخطوات تقف عليها معالجة مواقع الويب: -

أولاً : تحديد أي مواد مواقع الويب هي الأهم في التعبير عن محتواها أو موضوعاتها.

ثانيًا : الوقوف على المصطلحات الدالة في التعبير عن تلك المواد.

ثالثًا : اتباع ملامح إستراتيجية التنظيم التي يعتمد عليها في تنظيم مواقع الويب الحالية والمصادر المستقبلية.

---

(١) Cohen , Laura. Understanding the World Wide Web. Internet tutorials, 2007, cited 20/6/2007, cited at http://www.internettutorials.net/

تنتهي عملية تنظيم محتوى مواقع الويب عند قائمة بالمصطلحات الموضوعية مرتبة من العام إلى الخاص فالأكثر تخصصًا داخل أدلة بحث الويب، وقد ساعد على إخراجها بهذا الشكل مجموعة من العناصر البشرية القائمة على عملية تحليل محتوى مواقع الويب. أيضًا توجد نفس النهاية مع الاختلاف في الشكل داخل محركات بحث الويب ؛ حيث تنظم مواقع الويب داخل قواعد بيانات المحركات بالاعتماد على الكلمات الدالة وخصائص الكشاف المقلوب inverted index ، وقد جاءت هذه النهاية بأداء برنامج التكشيف داخل محرك البحث.

٤ / ٢ / ٢ / ٢ / ٣ تنظيم المحتوى التفاعلي الديناميكي لمواقع الويب:

كما سبق ذكره، فان مواقع الويب تتميز بسرعة تغيير المحتوى والتحديث الدائم. وكان لذلك أثره على تنظيم محتوى مواقع الويب ؛ حيث خرجت بعض الأفكار التي تدعو إلى استخدام خرائط مواقع الويب web sites maps ، ويتجلى دور هذه الخرائط في رسم الواقع الحقيقي والعملي لموقع الويب من خلال متابعة شبكة المسارات داخل بنية الموقع Site URLs structure. يعكس ذلك طبيعة ترابط صفحات الويب واتجاهاتها ، أيضًا متابعة نشاط أو انكسار بعض مسارات الموقع في عمليات التحديث المختلفة. وباختلاف فترات رسم هذه الخرائط لمواقع الويب يختلف التعبير عن شكل المحتوى ومن ثم اختلاف المصطلحات الكشفية الدالة على المحتوى. وتوفر البيانات الناتجة من وصف المحتوى مجموعة من التسجيلات records المترابطة عن محتوى كل صفحة داخل موقع الويب. ويعد الترابط بين التسجيلات باستخدام الروابط

الفائقة هو أهم الصفات اللازم توافرها في المعالجة الفنية لمواقع الويب. ويمكن أن تتسم هذه التسجيلات ببعض الضوابط مثل:

- إظهار هذه التسجيلات للمشتركين أو المسجلين داخل قائمة موقع الويب.
- إخفاء هذه التسجيلات بعيدًا عن أعين مستخدمي الويب مع الاحتفاظ بإمكانية إظهارها عند الحاجة.
- الإخفاء التام أمام مستخدمي الويب وإظهارها فقط أمام مسئولي الموقع واحدة من خصائص المسؤولية عن موقع الويب.

٤ / ٢ / ٢ / ٣ معالجة وتنظيم قواعد بيانات الويب:

تختلف قواعد بيانات الويب عن غيرها من مصادر الويب في المعالجة الفنية والتنظيم. حيث تعالج صفحات ومواقع الويب المعالجة الخارجية التي تشمل الحديث عن وصف وتنظيم الموقع كاملا كوحدة واحدة. غير أن المعالجة الفنية لقواعد البيانات تهتم بالمعالجة الداخلية لملفاتها واتباع طرق التنظيم الملائمة لطبيعة ملفات كل قاعدة بيانات على حدة. بشكل آخر أن المعالجة الفنية لمواقع وصفحات الويب تهدف إلى تنظيم المواقع معًا ، بينما تهدف المعالجة الفنية لقواعد البيانات إلى تنظيم مجموعة ملفات كل قاعدة وليس مجموعة قواعد بيانات الويب معًا غير أن هذا الواقع تغير الآن بالحديث عن تنظيم قواعد بيانات الويب معًا من خلال أدوات تنظيم وبحث تتعامل مع قواعد البيانات على مستوى قاعدة البيانات كوحدة وليس على مستوى الملف كما ألِفَ مستخدمو الويب منذ نشأة الويب. وأصبح الاتجاه نحو تقنين المعايير

الملائمة لوصف قاعدة البيانات كمصدر للمعلومات على الويب، وهو ما يعد من المشكلات الفنية الآن التي تشغل مصممي أدوات بحث الويب. ولهذا جاءت دراستنا الحالية.

٤ / ٣ أنماط نظم استرجاع الويب :

٤ / ٣ / ١ نظم استرجاع صفحات الويب :

حددت Laura Cohen أساليب استرجاع صفحات الويب في أربعة أشكال هي[١]:

١- استخدام مسار صفحة الويب URL داخل فراغ العنوان متصفح الويب.

٢- استخدام أسلوب التصفح اعتمادًا على الروابط بين صفحات الويب.

٣- البحث داخل إحدى القوائم الموضوعية لأحد أدلة بحث الويب المرتبطة بمجموعات الصفحات.

٤- استخدام فراغ البحث داخل أحد محركات بحث الويب بإدخال كلمات مفتاحية لاسترجاع المطابق لها من صفحات الويب.

ويمكن تناول مختلف نظم استرجاع صفحات الويب السابقة كما يلي:

٤ / ٣ / ١ / ١ نظام الاسترجاع باستخدام URL:

تعتمد الويب في بنائها على استخدام الروابط الفائقة للربط بين صفحات الموقع الواحد والمواقع المختلفة. ولذلك فإن أي ملف أو صفحة أو مصدر

---

(١) Cohen , Laura. Lbid.

معلومات على الويب لابد أن يحمل مسارًا أو عنوانًا ثابتًا يختلف لا يتطابق مع أي مصدر آخر على الإطلاق. والتسمية الشائعة لهذا العنوان هي URL أو محدد المصادر الموحد. وتخصص برامج تصفح الويب على اختلافها مكانًا للبحث بإدخال مسار المصدر يسمى Address Bar أو شريط العنوان. ولكي تتعرف برامج التصفح إلى هذا المصدر واسترجاعه أمام مستخدمي الويب ، فان هناك صيغة موحدة ثابتة لكتابة مسار ملف أو مصدر الويب تعتمد على أركان ثابتة ، يوضحها الجدول رقم (٤ / ٤) ومن ثم يستطيع أي مستخدم استدعاء موقع الويب بشكل محدد عند كتابة المسار بشكل صحيح. غير أن هذا الأسلوب من الاسترجاع يتطلب من مستخدمي الويب حفظ مسارات مواقع الويب موضع الاهتمام وهو ما لا يمكن لأي مستخدم القيام به.

الجدول رقم (٤ / ٤) مكونات مسار مواقع الويب

الوظيفة	عناصر المسار
بروتوكول النص الفائق	Protocol: **http**
نوع خدمة الويب والخادم المضيف للموقع	Host computer name: **www**
اسم موقع الويب	Second-level domain name:
نوع وطبيعة موقع الويب	Top-level domain name:
امتداد اسم الموقع	Directory name:
اسم الملف	File name:

٤ / ٣ / ١ / ٢ نظام الاسترجاع باستخدام برنامج التصفح: Browser

تشكل صفحات ومواقع الويب معًا نسيجًا عنكبوتيًا متداخل ومتقاطع المسارات. حيث يمكن نظريا تصفح كل صفحات الويب ومساراتها ؛ انطلاقًا

من أي صفحة ويب تشكل نقطة البداية للتنقل والتصفح داخل الويب. وليس من المبالغة القول إن هذه المسارات المتداخلة تشكل معًا النسيج العصبي لجسد الويب، الذي إذا ما فقدته الويب فقدت الترابط والقدرة على التنظيم. ولا يتطلب نظام الاسترجاع بتصفح مواقع الويب سوى استخدام أحد برامج التصفح مثل: internet explorer or Netscape navigator. غير أن هذا الأسلوب يأخذ مستخدمي الويب إلى أبعد مما يحتاجون من المعلومات المحددة، ذلك عن طريق ما يحدثه التصفح من تشتيت لأفكار الباحثين بمرورهم بصفحات تقترب أو تبعد من موضوع البحث. وتقسم برامج تصفح الويب إلى نوعين هما:

- **متصفحات رسومية: Graphical**

وهي البرامج التي تستطيع قراءة وسائط النص والصورة والصوت والفيديو. وأشهر برامج التصفح الرسومية؛ Internet Explorer, Netscape, Mozilla and Opera. وتعمل هذه البرامج في بيئة التشغيل الويندوز windows حيث يتم تحميلها بتحميل نظام تشغيل الحاسبات الشخصية.

- **متصفحات نصية: Text**

وتتصفح هذه البرامج الويب بشكل نصي يعرض فقط أسماء الملفات مع إظهارها بشكل مختلف وسط البنية العامة لصفحة الويب. ويتم التحكم في تصفح الويب باستخدام مؤشرات لوحة المفاتيح لأعلى وأسفل. وأشهر هذه البرامج ؛ Lynx.

٤ / ٣ / ١ / ٣ نظام الاسترجاع باستخدام أدلة بحث الويب:

تسمى أدلة بحث الويب أيضًا شجرة الموضوعات subject tree. وتعتبر أدلة الويب فهارس لمعالجة وتنظيم صفحات ومواقع الويب باستخدام العنصر

البشري، ومن الممكن أن تبدأ أدلة البحث بمجموعة قليلة من رؤوس الموضوعات ثم تتزايد بتزايد أعداد الصفحات والمواقع المضافة إلى قاعدة دليلي البحث. وتنظم صفحات الويب بحسب تحليل المحتوى والتعرف إلى الموضوعات التي تحويها بداخلها، ومن ثم تأخذ كل صفحة ويب المصطلح الدال على هذا الموضوع مع وضع رابط بينها وبين المصطلح الدال في قائمة رؤوس الموضوعات داخل دليل البحث. وما على مستخدم الويب سوى تصفح مصطلحات قوائم رؤوس الموضوعات داخل دليل البحث والضغط على رأس الموضوع مجال البحث لاستدعاء مختلف صفحات ومواقع الويب المرتبطة به. ويمتاز دليل البحث بدقة المواقع المستدعاة لموضوع محدد، وذلك يأتي من استخدام العنصر البشري في عملية المعالجة والتحليل الموضوعي لصفحات الويب. أيضًا فإن أدلة البحث تساعد مستخدمي الويب في الوصول إلى نتائج جيدة تتعلق بالموضوعات التي ليس لديهم فكرة كبيرة عنها، خلاف محركات البحث التي تتجه إلى التخصيص نتيجة استخدام الكلمات المفتاحية في عملية البحث. ومن أشهر أدلة البحث على الويب؛ دليل ياهو Yahoo الذي عمل مؤخرًا على إضافة آلية البحث بالكلمات المفتاحية كخيار إضافي لبحث قاعدة بياناته[١].

٤ / ٣ / ١ / ٤ نظام الاسترجاع باستخدام محركات بحث الويب:

يمكن وضع تعريف إجرائي لمحركات البحث هو " إحدى أدوات بحث استرجاع المعلومات على الويب، وتعتمد على ثلاثة برامج أساسية في ذلك هي الزاحف والمكشف وآليات البحث، بما يوفر القدرة على الإضافة والبحث.

(١) SUBJECT DIRECTORIES. Debbie Flanagan, 2004, cited 10/6/2007, cited at http://www.learnwebskills.com/search/subject.html.

"وتمثل محركات البحث العمود الفقري لتنظيم واسترجاع صفحات ومواقع الويب. حيث تعمل محركات البحث نظريًا على جمع وتنظيم وبحث كل صفحات ومواقع الويب التي تضاف كل يوم إلى بيئة الويب. وتختلف المحركات عن الأدلة كنظم استرجاع في أكثر من جانب مثل:[1]

أولاً : تختار أدلة البحث ما يضاف إليها بانتقاء العنصر البشري لطبيعة وتخصص دليل البحث، بينما يضم برنامج الزاحف في محرك البحث كل ما يواجهه من صفحات الويب بغض النظر عن طبيعتها أو محتواها.

ثانيًا : تستخدم أدلة البحث العنصر البشري في أداء عمليات المعالجة والتنظيم المختلفة، بينما يعتمد محرك البحث على برامج ثلاثة للقيام بالإضافة والتنظيم والبحث هي على الترتيب spiders or robots , indexer and search operators.

ثالثًا : تستخدم أدلة البحث حقول الوصف والتسجيلات الرقمية لمعالجة الصفحات والمواقع، بالإضافة إلى تنظيم الصفحات والمواقع أسفل قوائم رؤوس الموضوعات. وتعمل محركات البحث باستخدام المعالجة والتنظيم أسفل الكلمات المفتاحية.

رابعًا : تستخدم أدلة البحث تصفح قوائم الموضوعات وترتيبها من العام إلى الأخص لاسترجاع صفحات ومواقع الويب. أما محركات البحث فتستخدم الكلمات المفتاحية كأسلوب للاسترجاع.

ومن أشهر محركات البحث google, altavista and msn. وجدير بالذكر هنا الإشارة إلى تطور أدوات بحث الويب أدى إلى ظهور **البوابات** web

(١) سيد ربيع سيد. مصدر سابق.

gate way or portals. والبوابة هي أداة لبحث الويب تستخدم أسلوبي التصفح والكلمات المفتاحية في الاسترجاع. وتتميز البوابات عن الأدلة والمحركات بتقديم خدمات أخرى مثل : البريد الإلكتروني ومجموعات الأخبار وخدمات النقاش.

٤ / ٣ / ٢ نظم استرجاع مواقع الويب:

تتفق مواقع الويب مع صفحات الويب في اعتماد نظم الاسترجاع نفسها ، غير أن موقع الويب يزيد من صعوبة عمل وأداء نظم الاسترجاع السابقة. يميل موقع الويب المكون من مجموعة من الصفحات دائمًا إلى التغير في المحتوى وحجم الصفحات المترابطة داخله مما يزيد من صعوبة وصف المحتوى داخل الموقع سواء أكان ذلك على مدد زمنية يتخللها تحديث للموقع، أم كان على مستوى كل الصفحات الداخلية المكونة لموقع الويب. وتعجز بعض برامج الزاحف داخل محركات البحث عن الوصول إلى مستويات متعمقة؛ لتحليل موقع الويب المؤلف من عشرات الصفحات، نظرًا لمستوى التحليل المحدود في برنامج الزاحف.

يمكن القول إن مواقع الويب تقف على خط فاصل بين سهولة المعالجة لمحتوى صفحات الويب في نظم الاسترجاع، وبين قواعد البيانات التي تمثل المشكلة الرئيسة في المعالجة والتنظيم داخل نظم الاسترجاع على الويب. وهذا لأن موقع الويب يحمل نفس طبيعة بناء html مثل صفحات الويب، غير أنه يحمل طبيعة عمق المحتوى وديناميكيته التي توجد داخل قواعد البيانات. ولن يتم هنا تكرار ما سبق ذكره عن نظم استرجاع صفحات الويب نظرًا لتماثلها مع مواقع الويب ثابتة المحتوى.

٤ / ٣ / ٣ نظم استرجاع قواعد البيانات :

تختلف قواعد البيانات في طبيعتها عن بنية الويب؛ حيث تعتمد قواعد البيانات على نظم تشغيل مثل: MS SQL, Oracle, MS Access, My SQL. وهي بدورها لا تتعامل مع بنية النص الفائق في تنظيم ملفات قواعد البيانات. ويترتب على ذلك عدم استطاعة نظم استرجاع الويب الحالية التعامل مع محتوى هذه القواعد نظرًا لاختلاف طبيعتها عما نشأت له برامج الزاحف والمكشف. ولدخول هذه المصادر الرقمية (قواعد البيانات) إلى الويب تحتاج لبنية وسيطة تجمع بين خصائص واجهات النص الفائق من جانب وآليات عمل قواعد البيانات من جانب آخر. وسميت هذه الواجهات بـ CGI داخل خادم الويب من جانب المستخدم، بالإضافة إلى استخدام the JDBC interface من جانب قواعد البيانات. ويمثل ذلك الشكل رقم (٤-٧) (١).

إن ثمة مستويين للاسترجاع تمر بهما قواعد البيانات على الويب.

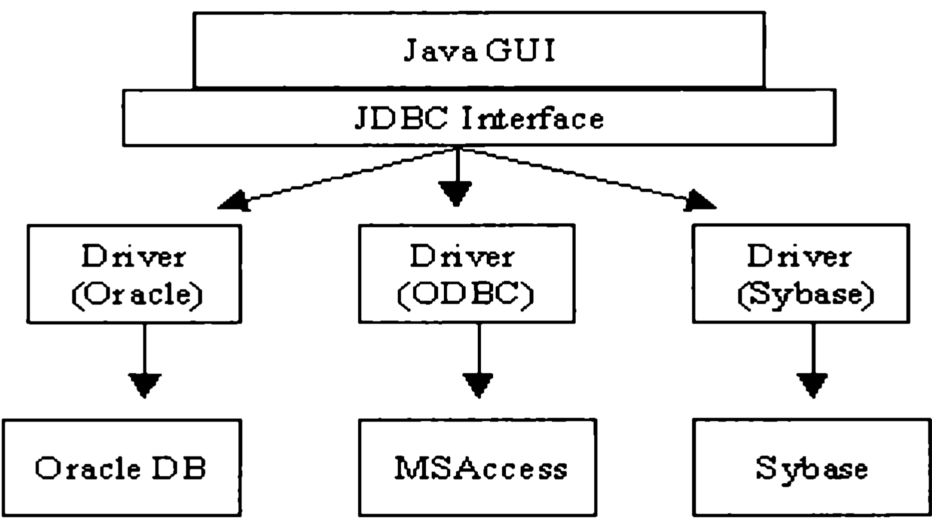
الأول : متمثلاً في أساليب استرجاع الملفات الداخلية المكونة لقاعدة البيانات ويقوم بذلك النظام الداخلي لإدارة قاعدة البيانات DBMS.

الثاني : متمثلاً في استرجاع قاعدة البيانات كوحدة واحدة ثم التعامل مع ما بداخلها من ملفات وتقوم بذلك محركات وأدلة بحث الويب غير المرئية. ويمثل الجانب الثاني المشكلة التي تواجهها نظم الاسترجاع الآن على الويب، ويطلق عليها ظاهرة الويب غير المرئية نظرًا لعدم تمكن هذه النظم من رؤية

---

(١) Sol, Selena. Introduction to Databases for the Web, 1998, cited 10/6/2007, cited at http://www.databasejournal.com/sqletc/article.php/26861_1431601_1

محتوى قواعد البيانات والقيام تجاهه بالمعالجة والتنظيم اللازم للاسترجاع أمام مستخدمي الويب.



الشكل رقم (٤ -٧) الواجهات الوسيطة بين الويب وقواعد البيانات

٤ / ٤ أشكال محتوى الويب:

تقسم محركات البحث محتوى الويب إلى قسمين ؛ الأول : الويب المرئية وهو كل ما تستطيع محركات البحث الوصول إليه ومعالجته وتنظيمه على اختلاف أنواع ملفات المعلومات أو وسائط حملها. الثاني: الويب غير المرئية وهو كل ما لا يمكن لمحركات البحث التعامل معه أو إضافته إلى مستودعاتها. فالقدرة على الوصول إلى المعلومات أو عدم الوصول إليها عند محركات البحث، تمثل قدرة المحركات على رؤية أو عدم رؤية المعلومات الرقمية. ومن ثم فإن معيار الحكم على المعلومات الرقمية بكونها مرئية أو غير مرئية داخل المحركات، هو وصول أو عدم وصول محركات البحث إليها. ويمكن توضيح شكلي المعلومات المرئية وغير المرئية كما يلي :

٤ / ٤ / ١ الويب المرئية:

يطلق على الويب المرئية مسميات أخرى مثل الويب السطحية surface web، وهو المعنى المقابل لأحد مصطلحات الويب غير المرئية المسمى بالويب العميقة the deep web. وقد أوردت موسوعة Wikipedia encyclopedia تعريفًا للويب المرئية هو "إن الويب السطحية – التي تعرف أيضًا بالويب المرئية أو المكشفة - هي ذلك الجزء من الويب القابل للوصول والمكشف بواسطة محركات البحث. وتكون محركات البحث مستودعاتها من مصادر الويب باستخدام برامج الزاحف أو العنكبوت التي تبدأ عملها بالاعتماد على قائمة مواقع معلومة مسبقًا، ويقوم الزاحف بالوصول لصفحات الويب ونسخها إلى مستودع محرك البحث. وتستخدم برامج الزاحف الروابط الحديثة في الوصول إلى صفحات أخرى، ومن ثم تكون كل صفحات الويب قابلة للوصول والتكشيف بواسطة محركات البحث. وفي حالة وجود عوائق تمنع برامج الزاحف من أداء عملها مثل كلمات المرور أو مستويات التحليل العميقة أو تقنيات متطورة ... فإن محرك البحث لا يرى هذه الصفحات وتدخل بذلك ضمن ما تعرف بالويب غير المرئية[١]."ويتضمن هذا التعريف بعض الجوانب مثل:

١- الربط بين مصطلحي الويب المرئية والويب المكشفة، بما يؤكد أن الرؤية إنما هي رؤية محركات البحث لصفحات الويب حتى يتثنى له التعامل معها.

---

(١) Surface Web. Wikipedia : the free encyclopedia, 2007, cited 20/6/2007, cited at http://en.wikipedia.org/wiki/ Surface_ Web.

٢- اشتمل التعريف السابق على كل جوانب عمل محرك البحث، حيث يكفي لأن تكون الصفحات مرئية أن يتم الوصول إليها بواسطة برنامج الزاحف أحد برامج عمل محركات البحث.

وتشمل الويب المرئية مختلف أشكال صفحات ومواقع وملفات والقليل من قواعد البيانات على الويب ؛ حيث تكون هذه المعلومات متاحة أمام برامج الزاحف للتعامل معها ونسخها داخل مستودعات محركات البحث. وتستخدم برامج الزحف تقنيات الروابط الفائقة بين صفحات الويب؛ لتشكل نسيجًا من مصادر المعلومات يظهر بعضه بعضًا.

٤ / ٤ / ٢ الويب غير المرئية:

يعرف قاموس مصطلحات المكتبات والمعلومات ODLIS الويب غير المرئية تحت مصطلح الويب العميقة على أنها " المعلومات المتاحة للوصول من خلال شبكة الويب العالمية، غير أنه لا يمكن استرجاعها بواسطة محركات البحث التي تعتمد على برامج الزاحف أو العنكبوت، مثل أشكال المعلومات التي توجد في ملفات pdf داخل قواعد البيانات التي تعتمد فقط على أسئلة البحث بشكل محدد. وتحوي الويب العميقة حجمًا من المعلومات يفوق الويب السطحية بـ ٤٠٠ -٥٠٠ مرة، يوجد أكثر من نصفها داخل قواعد البيانات، وتستخدم الويب بعض الخدمات لاسترجاع هذه المعلومات مثل CompletePlanet and ProFusion.[(١)] " ويظهر التعريف السابق بعض الجوانب مثل :

---

(١) ODLIS —Online Dictionary for Library and Information Science. OP.TC.

١- إن الأصل في المعلومات الرقمية هو الوصول من خلال محركات البحث التي تعد نافذة مستخدمي الويب لرؤية مصادر المعلومات الرقمية.

٢- إن السبب الرئيس لظهور الويب العميقة هو وجود المعلومات الرقمية في أشكال متنوعة مثل ملفات pdf , mdb, ppt....

٣- إن المصدر الرئيس لتكوين الويب غير المرئية هو قواعد البيانات، حيث تتصف بخصائص مخالفة لطبيعة العمل داخل محركات البحث بما يصعب على المحركات الوصول إلى محتوى تلك المصادر.

ويرى الباحث أن الويب المرئية تتكون من مجموعة من المصادر تقف في بنيتها الأساسية على بنية html التي تستخدمها محركات البحث للتعرف إلى المصادر حديثة الاقتناء، إلى جانب ما بداخل هذه الصفحات من ملفات ذات روابط محددة يمكن لمحركات البحث المتخصصة الوصول إليها بحسب نوع الملفات الذي تهتم به. أما الويب غير المرئية فإنها تعتمد على مكون أساس هو قواعد البيانات، وعلى ذلك لن يكون من الغريب أن يعتبر الباحث مصطلح قواعد البيانات بمثابة المترادف الموضوعي لمصطلحات الويب غير المرئية والويب العميقة والويب الخفية. وتمثل هذه الفكرة واقع العمل بين محركات البحث وقواعد البيانات ؛ حيث تستطيع محركات البحث الوصول إلى أي مصدر معلومات دون قواعد البيانات لما لها من طبيعة بناء وتصميم لا تتوافق مع محركات البحث في العمل ومن ثم يشكل محتوى هذه القواعد بيئة جديدة غير مرئية أمام محركات البحث.

وقد أورد الباحث في دراسته للماجستير قواعد بيانات الويب في مستويات عمق مختلفة، يمكن على أثرها تقسيم مستويات رؤية معلومات الويب بداية

من المعلومات ذات الرؤية غير الشفافة انتهاءً بالشكل غير المرئي تمامًا لقواعد البيانات. وكانت مستويات رؤية المعلومات كالتالي:[1]

٤ / ٤ / ٢ / ١ الويب غير الشفافة :

وهي تشمل قواعد البيانات كبيرة الحجم على الويب، التي تتكون بدورها من مجموعات كبيرة من مواد المعلومات، وعلى الرغم من ملاءمة هذه المواد لعمل برنامج الزاحف وإمكانية اقتنائها داخل قواعد محركات البحث، إلا أن برامج الزاحف لا تستطيع الوصول إليها ويرجع Paul Pedley هذا إلى عوامل مختلفة منها:

- عمق الزحف: تعمل محركات البحث على زيادة عمق التحليل والبحث لبرامج الزاحف؛ لتحقيق أفضل أشكال الاقتناء، خاصة في الوقت الذي تتزايد فيه أعداد صفحات مواقع الويب بشكل مستمر. وتعمق برنامج الزاحف داخل موقع الويب وتحليل كامل صفحاته لأجل جمعها من شأنه رفع تكاليف العمل داخل محركات البحث. وتتجنب محركات البحث إظهار مدى العمق الذي تعمل به برامج الزحف داخل صفحات الويب.
- تأخر التحديث الدوري: فإضافة صفحة جديدة إلى محرك البحث يحتاج إلى وقت قد يصل إلى شهر أو أكثر حتى يتم تمييزه ببرنامج الزاحف، وعلى الجانب الآخر، فإن الرجوع إلى مواقع سابقة للبحث

(١) سيد ربيع سيد. **محركات بحث الصور الثابتة : دراسة تحليلية** .- ط١.- الرياض : مكتبة الملك فهد الوطنية، ٢٠٠٧. -ص٣٥

عن التحديثات الأخيرة بها من صفحات مضافة يحتاج إلى وقت أطول من الوقت الذي يستغرقه التحديث . وعلى ذلك فإن جزءاً من مواد المعلومات المحدثة داخل مواقع الويب التي تم إضافتها مسبقًا سيظل بعيدًا عن أيدي الباحثين مدة من الوقت، وهو ما يحدث في النوع الحالي من قواعد البيانات.

- ازدياد النتائج المستدعاة : Maximum Number of Viewable Results تتسم محركات البحث بكبر حجم قواعد بياناتها ، ويتضح ذلك في أعداد النتائج المسترجعة داخل قاعدة البيانات التي قد يصل عددها إلى ملايين الصفحات، مما يعني معه أن محركات البحث تعمل على الإهمال المتعمد لاستبعاد النتائج الأقل علاقة بالنسبة لكلمات البحث ومثل هذه النتائج تظل دائمًا الجزء غير المرئي أمام الباحثين.
- زيادة الروابط غير النشطة: وهذا بدوره لا يرجع إلى عمل محرك البحث وإنما خطأ في تفعيل روابط بعض الصفحات التي لا يستطيع برنامج الزاحف الوصول إليها سوى باستخدام مسار الصفحة URL أو الإضافة المادية عن طريق مصممي صفحات الويب.

٤ / ٤ / ٢ / ٢ الويب الخاصة :

وهي تتكون من مجموعة من صفحات الويب القابلة للتكشيف بواسطة برنامج الزاحف، غير أنها تستخدم حالة من حالات ثلاث تالية لمنع دخول برنامج الزاحف إليها. وهذه الحالات هي:

- استخدام كلمات المرور للتعرف إلى محتوى الصفحة.

- تهيئة ملف robots.txt المحدد لعمل برنامج الزاحف لمنع التعامل مع صفحة الويب.
- استخدام حقل noindex meta tag للتوقف عن إضافة هذه الصفحة.

٤ / ٤ / ٢ / ٣ الويب ذات الملكية :

ويقصد بهذا النوع مجموعة قواعد البيانات التي تتوقف على وجود حساب أو ربط Registration مع موقع القاعدة. وعلى الرغم من أن هذا النوع يقدم خدمات مجانية، فضلاً عن سهولة إتمام عملية الربط بالقاعدة، إلا أن أفضل برامج الزاحف حتى الآن لا تستطيع التعامل مع إقامة ربط لقواعد البيانات. مما يجعل من الصعب عبور الزاحف إلى محتوى هذه القواعد.

٤ / ٤ / ٢ / ٤ الويب غير المرئية :

ويعد هذا الجزء من الويب هو الواقع الفعلي لمفهوم الويب غير المرئية. ويمكن لأي جزء من صفحات الويب أن ينجلي أمام الباحثين إذا ما امتدت إليه برامج محركات البحث لتكشيفه وإخضاعه للبحث أمام المستفيدين، ويختلف الجزء الحالي من الويب في البناء الفني عن الشكل المستخدم لدى برامج الزاحف، فلا يستطيع برنامج الزاحف استخدام محتوى تلك الصفحات. ويشير Paul Pedley إلى أن ذلك يتوقف على احتواء صفحات الويب على أنواع من الملفات لا تتعامل معها محركات البحث مثل الملفات المضغوطة وملفات pdf وملفات الفلاش Flash، ولا يكفي الجزء البسيط من حقول الميتاداتا المصاحب لهذه الملفات في إعانة برامج الزاحف على تحليلها وتكشيفها التي تحتاج إلى التعامل مع كامل المتن لصفحة الويب في تكشيفها وبحثها.

وقد حددت[1] Laura Cohen فئات عامة لمصادر الويب غير المرئية ، وهي التي تشكل الخصائص العامة لتكوين هذه الويب. وجاءت الفئات كما يلي:

الفئة الأولى: محتوى قواعد البيانات: يتكون محتوى قواعد البيانات على الويب من معلومات تم اختزانها داخل جداول صممت باستخدام برامج مثل Access, Oracle, SQL Server and DB2 ، ويتم الوصول إلى هذه المعلومات باستخدام استفسارات البحث الملائمة لبنية قاعدة البيانات. وعلى ذلك فمحتوى قواعد البيانات يختلف عن محتوى صفحات الويب الذي يتم الوصول إليه مباشرة.

الفئة الثانية: ملفات الويب غير النصية: مثل ملفات الوسائط المتعددة أو الرسومية أو البرامج أو تلك الملفات في شكل pdf , mdb, ppt and doc

الفئة الثالثة: المحتوى محدد الوصول: - وهي مجموعة المواقع التي تم حمايتها باستخدام كلمات مرور أو برامج حماية نظرًا لطبيعة المحتوى الحساسة. أو تلك الصفحات التي تتيح محتواها فقط من خلال استمارات الاشتراك لمستخدمي الويب.

الفئة الرابعة: المحتوى التفاعلي (الديناميكي) : وهي صفحات الويب التي تضع استمارات إدخال للمعلومات وتتغير بياناتها بواسطة مستخدمي الموقع.

الفئة الخامسة: الصفحات فاقدة الروابط : Unlinked content - pages حيث يسبب افتقاد هذه الصفحات إلى روابط فائقة مع الصفحات الأخرى قطع طريق الوصول أمام برامج الزاحف.

---

(١) Cohen, Laura. The Deep. internet tutorials, 2006, cited 20/6/2007, cita at http://www.internettutorials.net/deepweb.html.

الفئة السادسة: صفحات ذات تقنيات محددة: Scripted content - pages
حيث يتم الوصول إلى هذه الصفحات بواسطة روابط رسومية قد لا يتمكن برنامج الزاحف من التعامل معها.

إن مصطلح الويب غير المرئية قد نال اهتمامًا بالغًا في أدبيات الإنتاج الفكري حول بيئة الويب، وعلى الرغم من ذلك فإن بعض الباحثين مثل[1] Laura Cohen يرون أن مصطلح "invisible Web" يعاني من فقر في الدلالة والتعبير عن واقع الويب. وذلك يرجع في نظرهم إلى الأسباب التالية: -

١- يعبر مصطلح "invisible Web" فقط عن آلية عمل محركات البحث، حيث يفترض المصطلح أن كل ما تصل إليه محركات البحث من مصادر الويب هو الويب المرئية، وإن غير المرئي من المصادر هو ما لا تستطيع الوصول إليه وتكشيفه داخل مستودعاتها. ومن ثم افتراض أن مستخدمي الويب يستطيعون فقط استرجاع معلومات الويب من خلال محرك البحث، ولا أمل لهم في استرجاع معلومات أخرى خارج محرك البحث، إلا أن يؤدي الحظ دوره في ذلك.

٢- لا يمكن القول إن هناك معلومات رقمية مسجلة يطلق عليها غير مرئية، وإنما يمكن القول إن هناك معلومات أصعب بدرجة ما في الحصول عليها من المعلومات الأخرى.

٣- إن قواعد البيانات على الإنترنت موجودة منذ وقت طويل، مما جعل مستخدمي الويب على قدر من التمرس بالتعامل وطرح الاستفسارات البحثية المباشرة لاسترجاع محتوى هذه القواعد.

(١) Cohen, Laura. The Deep.OP.TC.

٤- إن محتوى محركات البحث ذاته يتم اختزانه في قواعد بيانات، وهذا لا يشكل بطبيعته جزءاً من الويب غير المرئية؟ حيث لا يستطيع مستخدم الويب استرجاع ما بداخل محرك البحث سوى باللجوء إلى محرك البحث وطرح الاستفسارات البحثية داخل واجهة البحث.

غير أن الباحث يتبنى رؤية مخالفة تمامًا ترد على الجوانب السابقة ؛ حيث تجاهلت الأسباب السابقة حقيقة أن محركات البحث هي واسطة العقد في عمل الويب بين منتج لصفحة الويب ومستخدم لها : فمن غير الممكن أبدًا أن يحفظ مستخدمو الويب جميع مسارات صفحات ومواقع وقواعد بيانات الويب. هذا إلى جانب أن عددًا قليلاً من محركات البحث يمكنه نظريًّا تغطية كل المعلومات الرقمية الواقعة في فضاء الويب آخذين في الاعتبار التقدم والتطوير الذي تشهده محركات البحث.

إن ثمة حقيقة يجب التعرض لها وهي الربط بين قواعد البيانات عامة وبين الويب غير المرئية. فمصطلح الويب غير المرئية يطلق فقط على قواعد البيانات التي تعمل في بيئة الويب html, xml، وذلك لأن هناك الكثير من قواعد البيانات التي تعمل على الخط المباشر والفارق جلي بين النوعين ؛ حيث تحمل قاعدة بيانات الويب خصائص خلاف ما توجد عليه قواعد الخط المباشر ويتضح ذلك من أساليب الاسترجاع والعمل لكل منهما. تختلف بيئة الويب عن الخط المباشر فيما يلي(١):

---

(١) online DB and searching.2004, cited 15/7/2007, cited at http://web.simmons.edu/~benoit/LIS530/602slides/17online.htm.

- يعني مصطلح الخط المباشر تلك القواعد المتاحة على أقراص ضوئية CDs أو قاعدة محلية الاستخدام أو فهرس متاح على شبكة داخلية للمكتبة OPAC.
- ترتبط قواعد بيانات الخط المباشر بالتعامل مع الشكل العلائقي لقواعد البيانات RDBMS.
- يعني مصطلح الويب تلك الملفات المعالجة في بيئة html, xml ، التي يتم الوصول إليها عن بعد أو المختزنة في قواعد بيانات الويب.
- يعمل الاسترجاع على الخط المباشر بخصائص البحث والمضاهاة لمجموعات البيانات الببليوجرافية اعتمادًا على آليات البحث البولياني. أما الاسترجاع على الويب فغالبًا ما يعمد إلى التكشيف والاسترجاع للنص الكامل، وهو ما تقوم به محركات بحث الويب.
- تعمل بيئة الويب كبيئة تفاعلية دائمة التغيير والإضافة والتحديث، خلاف ما يوجد عليه الخط المباشر من ثبات أو بطء في التغيير.

ويختتم الباحث التعرف إلى الويب بمجموعة من الخصائص التي تظهر طبيعة الويب كنظام استرجاع في مقابل نظم استرجاع المعلومات الأخرى. وتتمثل هذه المقارنة في الجدول التالي:

الجدول رقم (٤-٥) خصائص الويب كنظام استرجاع مقابل نظم استرجاع المعلومات الأخرى

وجه المقارنة		The web	Information retrieval
الوثائق documents	اللغة languages	تتعدد لغات الوثائق داخل الويب، وتستخدم محركات البحث كلمات النص للتكشيف، ولا تستخدم أساليب تنظيم أخرى.	تتجه إلى التوحد في لغة الوثائق داخل قواعد بياناتها ، يتم ترجمة لغات النصوص الأخرى إلى نفس لغة قاعدة البيانات.
	نوع الملف file type	تتعدد أنواع الملفات بها ، قد تتسم بالصعوبة في التكشيف لغياب النص المصاحب.	عادة ما تقتني ملفات ذات شكل موحد ، وقد يكون الاعتماد على التسجيلات الببليوجرافية المصاحبة.
	طول الوثائق document lenght	تتنوع الوثائق بين بالغة القصر ومتناهية الطول، التي قد تقسم على أجزاء مثل صفحات موقع الويب.	قد تختلف الوثائق في طولها غير أنها جميعًا تعرض على النظام كوحدة واحدة.
	بنية الوثيقة document structure	تعتمد على بنية HTML التي قد تسمح بالبحث بشكل يسير.	تعتمد على بنية معقدة، بما يستلزم معه صياغة حقول بحث معقدة.
	وثائق الشوشرة spam	تحتاج محركات البحث هنا للوقوف على الوثائق المستحقة للتكشيف، دون تلك التي لا تمثل أي قيمة أو أهمية.	يستخدم الوثائق ذات الفائدة والأهمية منذ بناء قاعدة البيانات.

	الروابط الفائقة hyper links	ترتبط وثائق النظام من خلال كم هائل من الروابط الذي قد يحدد منزلة الوثيقة فيما بعد أثناء عمليات الاسترجاع.	لا ترتبط الوثائق معا داخل قاعدة البيانات، وتستخدم الاستشهادات الواردة داخل الوثيقة لتحديد قيمة المعلومات.
خصائص الويب Web characteristic	حجم البيانات وحجم قواعد البيانات Mount of data and size of databases	لا يمكن بحال تحديد الحجم الحقيقي لمعلومات الويب، ولا يمكن الوصول إلى وقت يكتمل في تكشيف كل وثائقها.	يتم استخدام أساليب معالجة وإحصاء لتحديد حجم المعلومات وكم الوثائق داخل النظام، ولابد أن تكون جميعها مكشفة في وقت واحد.
	التغطية coverage	التغطية غير معروفة، غير أنه يمكن تقديرها.	تتحدد التغطية تبعا لحجم وكم المعلومات والوثائق في الموضوع.
	التكرار duplicates	تتكرر العديد من الوثائق في إصدارات ونسخ مختلفة.	لا يحدث تكرار بين الوثائق حتى بين الوثائق ذات الطبعات المختلفة.
سلوك المستفيد User behavior	رغبات المستفيدين user intersets	تتعدد وتتباين رغبات واحتياجات المستفيدين من المعلومات بعدد من يستخدم الويب على مستوى العالم.	دائما ما يتحدد سلوك واحتياجات المستفيدين من المعلومات على مجموعة محددة، نظرا لاتجاه كل نظم الاسترجاع إلى الاقتصار على مجال موضوعي محددة. وبالتالي يقتصر استخدامه
	نوعية الاستفسارات Type of queries	يستخدم المستفيدون من كلمتين إلى ثلاثة للبحث نظرا لمحدودية المعرفة بالويب.	يتمرس المستفيدون بلغة البحث نظرا لثباتها واستخدامها بشكل موحد داخل النظام.

كنظام استرجاع للمعلومات IR system	واجهات البحث User interfaces	تستخدم الويب واجهات بسيطة سهلة الاستخدام من جانب المستفيدين.	عادة ما تكون واجهات معقدة تحتاج إلى التمرس.
	الترتيب والتنظيم Ranking	يتم الترتيب بحسب درجة الصلة بين الوثائق ومحتواها، وبين استفسار البحث وموضوعه.	عادة لا يستخدم النظام الترتيب ذات الصلة نظرا لتعود المستفيدين على استرجاع وثائق محددة.
	مهام البحث Search functions	تستخدم الويب أساليب واستراتيجيات بحث بسيطة.	عادة ما تعتمد على لغات الاستفسارات المعقدة.

٤ / ٥ الخاتمة:

تحتوي الويب على كم هائل من المعلومات الرقمية تختلف في أشكالها وملفاتها ووسائط حملها. وتحفظ الويب هذه المعلومات في صفحة أو موقع أو قاعدة بيانات على الويب. وتتبنى الويب بناء مصادر معلوماتها بالاعتماد على لغات الترميز المعيارية التي تتوافق من ناحية أخرى مع آليات عمل أدوات البحث على الويب من محركات وأدلة بحث. . ولكل من صفحات ومواقع وقواعد البيانات خصائص محددة في المعالجة الفنية والتنظيم للمعلومات وآليات محددة للاسترجاع. غير أن هناك مصادر معلومات مثل قواعد البيانات وبعض المصادر الأخرى قد شكلت بخصائص بنائها معلومات بعيدة عن متناول محركات البحث بما جعلها بعيدة عن متناول مستخدمي الويب.وعلى ذلك أضحت الويب منقسمة بين محتوى مرئي لمحركات البحث ومستخدميها، وبين محتوى آخر غير مرئي.

الفصل الخامس

الويب نماذج نظم استرجاع المعلومات

٥ / ٠ التمهيد:

تنطوي الويب كبيئة لتنظيم واسترجاع المعلومات الرقمية على نسيج متشابك من مصادر المعلومات ونظم الاسترجاع. لقد اتسمت الويب بملامح تختلف عنها في نظم استرجاع المعلومات التقليدية أو على الخط المباشر ؛ حيث تواجه الويب سيلاً جارفًا من المعلومات متعددة الأشكال ومتباينة الموضوعات ومتنوعة وسائط حمل المعلومات، وتتخذ من محركات بحثها وأدلتها جسرًا للنجاح في تنظيم واسترجاع مختلف المعلومات الرقمية التي تجملها الويب. وهذه الديناميكية والتفاعلية قد لا تواجهها بقية أنواع نظم الاسترجاع نظرًا لبطء التحديث أو لمحدودية المعلومات التي تتعامل معها هذه النظم. ويهدف هذا الفصل إلى دراسة نماذج لنظم استرجاع المعلومات على الويب من حيث تعاملها مع أشكال الوثائق المختلفة، وسبل المعالجة والتنظيم للوسائط المختلفة لحمل المعلومات؛ فضلا عن إلقاء الضوء على اختلاف مستويات العمل بين نظم استرجاع الويب ؛ حيث يختص بعضها بتناول المعلومات الرقمبة ذاتها، والبعض الآخر لتناول النظم العاملة في بيئة الويب.

٥ / ١ نظم استرجاع المعلومات على الويب:

تتسع بيئة الويب لتشتمل على الكثير من مواد المعلومات التي تقدم محتواها في شكل كلمات أو عناصر مرئية كالصور الثابتة والمتحركة أو في شكل صوت مسموع. ومن ثم عملت نظم استرجاع الويب على التخصص في معالجة وتنظيم أشكال محددة من مواد المعلومات، خاصة مع تنامي أعدادها وتنوع أشكالها، هذا فضلا عن الأشكال المستحدثة كالمواد ثلاثية الأبعاد

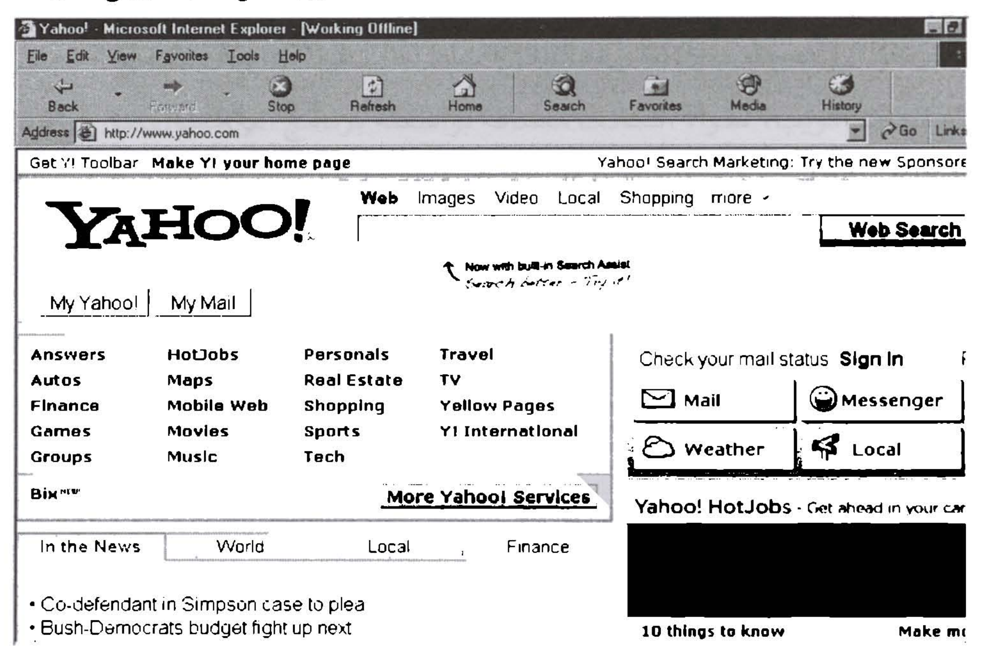
والرسوم الحيوية. وفيما يلي يقدم الباحث نماذج من نظم استرجاع المعلومات وقد اشتقت ترتيبها وتسلسلها من طبيعة المعلومات التي تحملها الويب :

٥ / ١ / ١ نظم استرجاع النصوص:

تتعامل نظم الاسترجاع اللاحقة مع النصوص كما تتعامل أيضًا مع مختلف أنواع الوسائط الأخرى، غير أن التركيز على النصوص هنا يعود إلى بداية ونشأة هذه النظم، كما أن غالبية هذه النظم يتعامل مع صفحات ومواقع الويب تنظيمًا واسترجاعًا. ومن أشهر نماذج نظم استرجاع النصوص الحالية على الويب النظم التالية:

٥ / ١ / ١ / ١ البوابات:

جاءت البوابات web portals نتيجة لاحتياج مستفيدي الويب لأداة تجمع بين ميزات كل من محركات وأدلة البحث. وهذا من ناحية بحث واسترجاع معلومات الويب، أما من ناحية الخدمات الأخرى، فإن بوابات الويب تقدم خدمات متقدمة لمستخدميها مثل : البريد الإلكتروني ومجموعات النقاش وقوائم البريد الإلكتروني... وغيرها. وتعد البوابات كنظام استرجاع على الويب نتيجة لاستجابة محركات وأدلة بحث الويب للتطور الذي تسير عليه ؛ حيث يوجد الكثير من البوابات الآن التي بدأت كمحرك أو كدليل بحث، وخير أمثلتها بوابة الويب ياهو !Yahoo. وإضافة لجوانب التميز أيضًا تعمل بوابات الويب على بحث مختلف أشكال المعلومات الأولية إلى جانب معالجة وبحث نظم استرجاع الويب كقواعد بيانات الويب، أي أنها تؤدي نفس دور محركات البحث المتعددة.



الشكل رقم (٥ –١) واجهة استخدام بوابة الويب (Yahoo!) والخدمات المقدمة منها.

إن بوابات الويب كنظام استرجاع منوطة بالتعامل مع مجموعات مصادر معلومات الويب على اختلاف مستويات هذه المعلومات واختلاف أشكالها، ومن هذه المصادر ما يلي(١):

١- الفهارس والبيانات الخلفية (الميتاداتا) والمصادر الببليوجرافية المتاحة من خلال بروتوكول الإنترنت TCP/IP أو معيار البحث Z39.50.

٢- صفحات الويب المكتوبة بلغة HTML واللغات الأخرى المتاحة من خلال بروتوكول الإنترنت TCP/IP.

---

(١) Search Portals. GILS, 2003, cited 25/11/2007, cited at http://www.gils.net/portals.html.

٣- مختلف أنواع ملفات البيانات المتاحة من خلال بروتوكول الإنترنت وبروتوكول FTP.

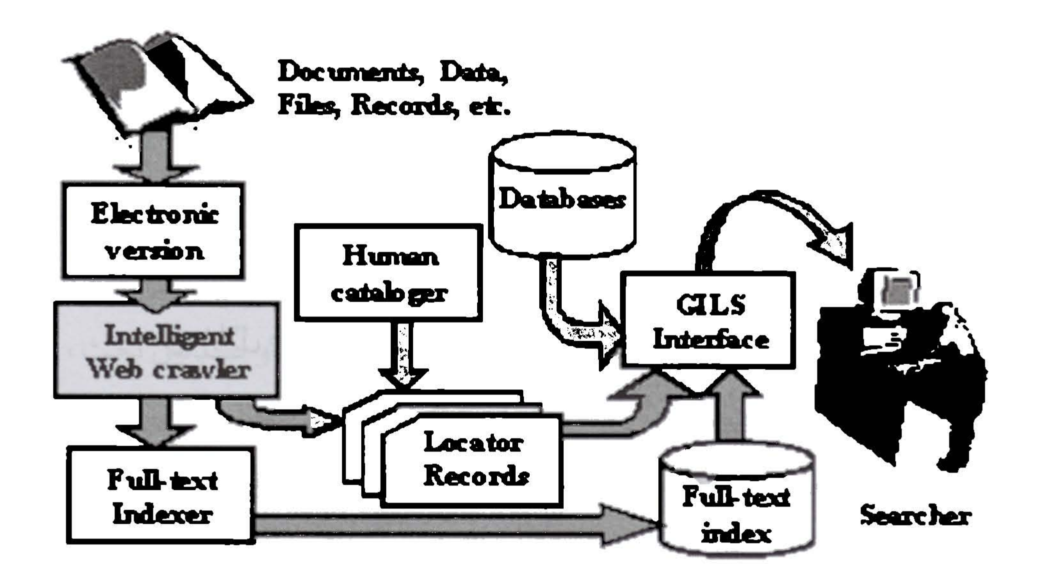
٤- مختلف أنواع المعلومات المنتجة بواسطة أنظمة المكتب office systems مثل Lotus Notes or Microsoft Office المتاحة من خلال شبكات المعلومات المحلية.

٥- مختلف قواعد البيانات مثل MS Access, MS SQL Server, Oracle المتاحة من خلال شبكات المعلومات المحلية أو بروتوكول الإنترنت TCP/IP.

٦- أدلة البحث المتواجدة في فضاء الويب.

٧- كشافات النص الكامل المتاحة من خلال بروتوكول الإنترنت.

ويصف الشكل رقم (٥ -٢) آليات عمل بوابات الويب وتفاعلها مع مصادر الويب الأخرى.



الشكل رقم (٥ -٢) آليات عمل بوابات الويب

تعمل البوابات كنظم استرجاع اعتمادًا على التزويد بالمعلومات من صفحات الويب، سواء أكانت مضافة بواسطة برنامج الزاحف أو من قواعد البيانات الأخرى. ثم تقوم البوابات بمعالجة وتنظيم هذه المعلومات لاسترجاعها بأنماط مختلفة مثل آليات البحث وأساليب التصفح المتاحة من خلال أدلة بحث الويب. ومن ثم يرى الباحث أن البوابات كنظم استرجاع للمعلومات تعكس التطور والتكامل بين أدوات بحث معلومات الويب، بما تحققه من شمولية وجمع بين مختلف مصادر المعلومات ومختلف آليات تنظيمها واسترجاعها.

٥ / ١ / ١ / ٢ أدلة البحث:

يطلق على هذا النوع من أدوات البحث مسميات كثيرة مثل subject guides internet resource guides , internet directories, internet indices , internet catalogs وكلها تعطي معنى القوائم أو الموضوعات المقسمة. وتعتمد أدلة البحث على تقسيم المعرفة أو العلوم إلى أقسام عامة ثم إلى أقسام أكثر تخصصًا أو قوائم فرعية، ومن ثم يمكن الحصول على الموضوع ومن ثم المواقع التي ترتبط مباشرة بهذا الموضوع المخصص. يعرف قاموس مصطلحات المكتبات والمعلومات على الخط المباشر ODLIS دليل البحث web directory على أنه " إحدى خدمات الخط المباشر التي تكشف مواقع ومصادر الويب المتاحة على الإنترنت، وهي تتيح الوصول إلى مصادر الويب من خلال اتباع نظام تصنيف هرمي للموضوعات، ومن أهم أدلة الويب دليل Yahoo." وطبقًا لذلك فإننا نعرف دليل الويب غير المرئية على أنه إحدى أدوات بحث مصادر الويب غير المرئية مثل قواعد البيانات وغيرها، ويعمل على استدعاء هذه المصادر باستخدام إحدى قوائم رؤوس الموضوعات أو خطط التصنيف المرتبة

منطقيًا. وتعتمد أدلة البحث على العنصر البشري في معالجة وتنظيم مصادر الويب، خلاف محركات البحث التي تعتمد كلية على برامج فرعية مثل الزاحف والمكشف وآليات البحث[1].

ولقد وضع Ross Tyner مجموعة من العناصر التي يتم عندها الاستعانة بأدلة البحث، إذا ما أراد الباحث الحصول على معلومات قيمة في مجاله وهي[2]:

١- عندما تكون هناك فكرة أو موضوع واضح في ذهن الباحث.

٢- عندما يريد الباحث حصرًا بالمواقع في موضوع واحد.

٣- عندما يريد الباحث الحصول على قائمة بالمواقع التي تخص موضوع مخصص لا يتخللها شوشرة من المواقع التي لا تتصل بالموضوع.

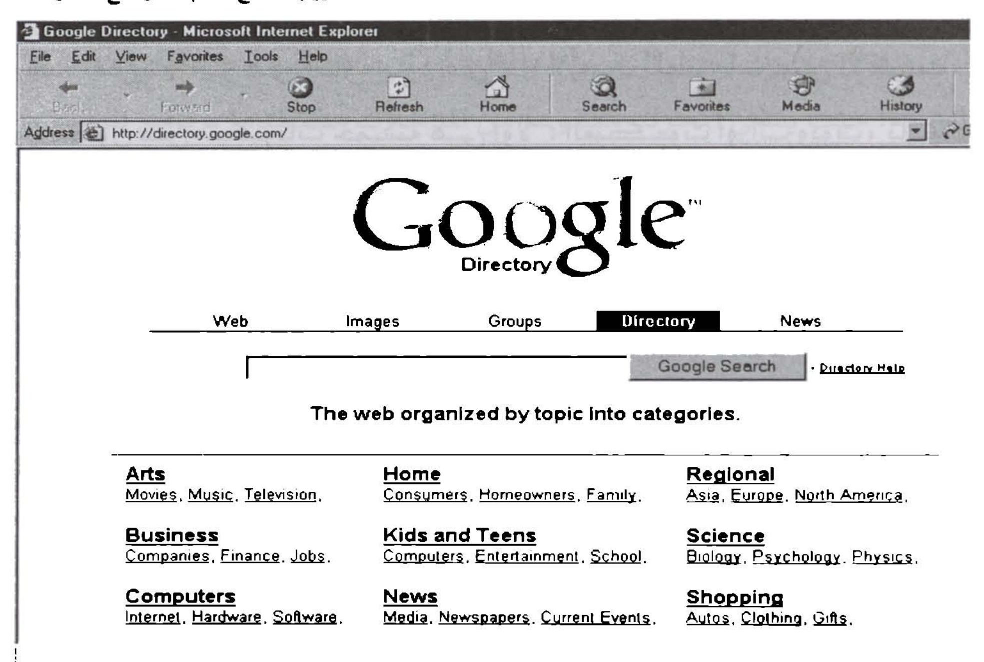
٤- عندما يريد الباحث الحصول على عناوين أو كلمات مفتاحية بغض النظر عن الحصول على مواقع أو مصادر تتصل بالموضوع.

٥- عندما يريد الباحث تجنب عرض تلك الصفحات التي تحتوي على محتوى قليل من الموضوع، والتي غالبًا ما تكون ضمن نتائج محركات البحث.

---

(١) Reitz, Joan M. directory, ODLIS — Online Dictionary for Library and Information Science, Libraries Unlimited, 2006, cited 20/8/2007, cited at http://lu.com/odlis/odlis_d.cfm#directory.

(٢) سيد ربيع سيد. **محركات بحث الصور الثابتة على الإنترنت**.- الرياض : مكتبة الملك فهد الوطنية، ٢٠٠٧.



الشكل رقم (٥ -٣)واجهة استخدام دليل البحث في جوجل (**Google**).

٥ / ١ / ١ / ٣ محركات البحث:

لقد ذكرت بعض مواقع الويب تعريفات عن محركات البحث منها ؛ " إن محرك البحث هو بصفة عامة أداة تمكن مستخدمي الويب من البحث بمصادرها ، وأنه يمكن أن يعتمد على البرامج أو الإنسان في عمله[١]"وهذا التعريف قد عمل على تلخيص مهمة محرك البحث وإبرازها من ناحية مستخدمي الويب دون التطرق إلى الإجراءات التي يقوم بها في تنفيذ هذه المهمة

(١) Sullivan , Danny . Search Engine Optimization & Marketing Glossary, Submitawebsite, Inc , 2004 , Visited at 12/6/2004 ,
http://www.submitawebsite.com/seo_glossary.html

وثمة تعريف آخر هو "أنه عبارة عن أداة لإيجاد المعلومات على الإنترنت، ويتكون من مكونات خمسة هي أولاً : العنكبوت أو الروبوت، ثانيًا: المكشف، ثالثًا : قاعدة البيانات، رابعًا : آليات البحث، خامسًا : واجهة الويب[1]". وقد اختلف هذا التعريف عما سبقه من ناحية التركيز على مكونات محرك البحث، وإن كان الباحث يرى أنها قد تقتصر فقط على ثلاثة مكونات هي الزاحف والمكشف والبحث. وعلى ما سبق من تعريفات يرى الباحث أن أفضل تعريف لمحركات البحث هو " إحدى أدوات بحث استرجاع المعلومات على الويب، وتعتمد على ثلاثة برامج أساسية في ذلك هي الزاحف والمكشف وآليات البحث، بما يوفر القدرة على الإضافة والبحث."[2]

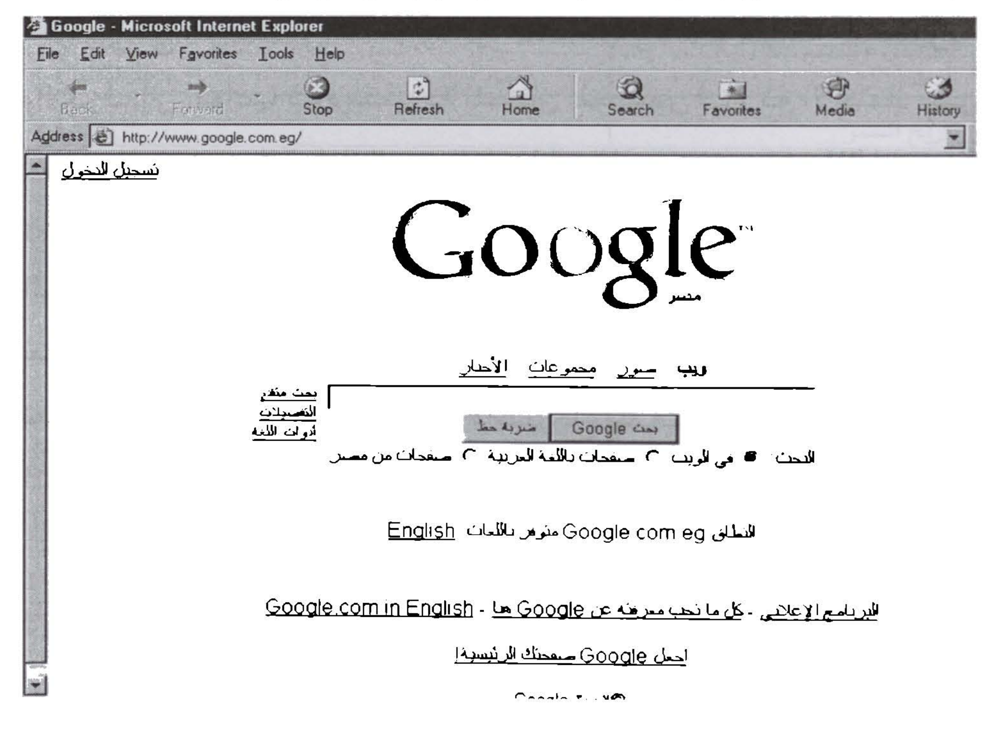
وتختلف محركات بحث الويب غير المرئية عن محركات الويب المعروفة في أنها تتعامل مع مصادر معلومات رقمية ذات بنية تختلف وبنية html. وأهم هذه المصادر اختلافاً في البنية والتنظيم هي قواعد البيانات موضع الحديث. يمكن اقتراح تعريف إجرائي لمحركات بحث الويب غير المرئية على أنها "إحدى أدوات بحث واسترجاع الويب، التي تختص بتنظيم نوع من مصادر الويب ذات بنية مختلفة عن بنية HTML ، ويمثل هذا النوع من مصادر الويب في أغلبه قواعد البيانات التي لا تستطيع محركات بحث الويب العادية الوصول إليها وتنظيمها أو استرجاعها."

---

(١) What is a Search Engine? Metamend.inc , 2004 , Visited at 12/6/2004 , http://www.metamend.com/search-engine-map.html

(٢) سيد ربيع سيد. مصدر سابق.

ويمثل محرك بحث الويب أهم نظم استرجاع صفحات وملفات معلومات الويب ؛ حيث يقتني المحرك مجموعة من المدخلات مستخدمًا برنامج الزاحف، تتحدد نوعية هذه المدخلات مسبقًا بحسب طبيعة تخصص محرك البحث كالاقتصار على صفحات الويب أو ملفات الصور أو ملفات الصوت ...الخ. وتتم معالجة وتنظيم معلومات الويب باستخدام تقنيات تكشيف النص الكامل مع الاعتماد على بعض حقول البيانات المصاغة بلغة HTML مثل حقول ؛ meta title, meta key word. وتختزن هذه المعلومات داخل مستودع محرك البحث مع ملف الكلمات الدالة inverted file. ويخصص محرك البحث مجموعة من الآليات لبحث واسترجاع محتوى مستودع المعلومات، باستخدام إستراتيجيات البحث المصاغة من جانب المستفيد ذاته والمعبرة عن موضوعه.



الشكل رقم (٥ –٤) واجهة استخدام محرك البحث (Google).

٥ / ١ / ٢ نظم استرجاع الوسائط المتعددة:

٥ / ١ / ٢ / ١ نظم متخصصة لاسترجاع الوسائط المتعددة:

٥ / ١ / ٢ / ١ / ١ محركات بحث الفيديو: Video Search Engines

تمثل محركات بحث الفيديو أحد أهم أنواع محركات بحث الوسائط المتعددة ، إلا أنه لا يوجد لهذا النوع من وسائط حمل المعلومات محركات بحث متخصصة في بحث ملفات الفيديو كما هو الحال في بحث الصور والمواد الصوتية. تعمل محركات بحث الفيديو بنفس المراحل التي تعمل بها محركات بحث النصوص والوسائط الأخرى، يقوم محرك البحث بالاعتماد على برنامج الزاحف أو العنكبوت في استقطاب والتقاط ملفات العنكبوتية والتعرف إلى أنواعها ثم وضعها في المكان المخصص لها، ثم الاعتماد على برنامج المكشف في تحليل النص التابع لمادة الفيديو؛ من مسار URL وعنوان وكلمات واصفة Meta Tags ، يخرج منها برنامج الزاحف بمجموعة الكلمات التي يستخدمها الباحثون في استدعاء ملف الفيديو. هذا من ناحية أداء المحرك نفسه، إلا أن الأمر الآخر هو ما يخص المحتوى نفسه، فان محاولات كثيرة تعمل جاهدة لإيجاد معايير لتحليل واسترجاع ملفات الفيديو سواء الوصف المادي أو الموضوعي.

وقد عمل معيار Dublin Core على التقعيد للوصف المادي والموضوعي لأوعية معلومات الفيديو، كما عمل على الصور من قبل ، ويأخذ الوصف هنا اتجاه التقسيم لتحليل موضوع هذه المادة إلى أقسام أو مقاطع ثم تقسيمها إلى مشاهد والمشاهد تقسم بدورها إلى لقطات ثابتة، يأتي ذلك في دراسة Jane

Hunter عن إقامة نظام تكشيف وتصفح واسترجاع يعمل في مكتبات المواد السمعية المرئية، مراعيًا أن يكون الوصف المادي لمواد الفيديو الرقمية شاملاً كل قسم من الأقسام المكونة لمادة الفيديو. وقد كان الوصف الكامل لوعاء الفيديو الرقمي تبعًا لمعيار Dublin Core كما يلي[1] :

الجدول رقم (٥-١) يوضح عناصر الوصف المادي لمواد الفيديو الرقمية على الإنترنت

النص	البيانات " الحقول " Meta data
عنوان المادة الفيلمية	Title =
القائم على إنتاج المادة الفيلمية	Creator = Producer
موضوع المادة في كلمات محدودة	Subject =
وصف الموضوع في عبارات شارحة	Description =
الناشر	Publisher =
تاريخ النشر	Date =
نوع المادة الفيلمية	Type = "Image.Moving.Film.Documentary"
وصف الشكل المادي " مقابل حقل التوريق "	Format = 1 videocassette (27 min.) : sd., col. ; 1/2 in
محدد الحصول على المادة	Identifier =
مصدر المادة الرقمية	Source =
اللغة	Language =
الأجزاء المكونة للمادة الفيلمية، الرقمية	Relation.HasPart = scene1, scene2, scene3, scene4, scene5,...
التغطية للمادة " الزمنية والمكانية "	Coverage =

(١) Hunter , Jane and Newmarch, Jan . An Indexing, Browsing, Search and Retrieval System for Audiovisual Libraries , Uni. of Queensland , 2001 , cited 1/10/2002.

إلا أن التحليل الموضوعي لمواد الفيديو ما برح يقف في منطقة ضبابية، تحتاج إلى الكثير من الجهد للكشف عن الطرق الفعالة في تكشيف وتحليل موضوع مثل هذا النوع من الوسائط المتعددة .

٥ / ١ / ٢ / ١ / ٢ محركات بحث الصور: Search Engines Image

تتفوق محركات بحث الصور عن محركات بحث الوسائط الأخرى في التواجد العددي على العنكبوتية، كما أن محركات بحث الفيديو يعتريها التخصص في التعامل مع ملفات الصور سواء تلك العامة " الطبيعية " أو الإعلامية أو الصور العلمية. تتميز أدوات العنكبوتية المتخصصة في بحث الصور عن الوسائط الأخرى بأنها تتكون من فئتين متباينتين في العمل هما ؛ محركات بحث الصور، وقواعد بيانات الصور. أن كلا النوعين يختلف عن الآخر في طبيعة التعامل مع الصور من ناحية التحليل المادي والموضوعي، أيضًا فان طبيعة التعامل في كل نوع تختلف عن الآخر بطبيعة مفهوم محرك البحث الذي يعمل بطريقة آلية خالصة تعتمد على ثلاثة برامج في إضافة وتنظيم وبحث ملفات العنكبوتية على اختلاف أنواعها الموضوعية والشكلية، خلاف ما توجد عليه قواعد البيانات من التدخل البشري في معظم إجراءاتها مثل تحديث قاعدة البيانات بإضافة المعلومات إليها ، وتنظيم مجموعات الملفات التي توجد فيها ، خلاف مرحلة البحث فإنها على غرار محركات البحث توكل بها إستراتيجيات بحث تتكون من آليات بحث مختلفة .

تتساوى الصور الرقمية على العنكبوتية مع أنواع الوسائط الأخرى في الحاجة إلى التوصيف المادي والموضوعي، وهو أحد محاور دراسة الباحث في

أطروحة الماجستير[1]، ومن الممكن أن تكون الصور الرقمية هنا مثالاً لتناول التحليل المادي والموضوعي للوسائط الأخرى ؛ حيث إن جميع ملفات الصور الرقمية على العنكبوتية مع اختلاف موضوعاتها " العلمية والطبيعية والإعلامية " تتساوى في الوصف المادي للخصائص الرقمية وحقول الترميز داخل ملفات لغة الترميز المعيارية HTML ، إلا أن التحليل الموضوعي يختلف من نوع إلى آخر من الصور. بالمثل فإن الباحث يرى أنه على مستوى كل نوع من الوسائط ؛ الفيديو أو الصوت فإن ملفات هذا النوع تتساوى في الخصائص الرقمية، مما يمكن معه إيجاد معايير لوصف وتحليل هذه الملفات. إلا أن الصعوبة في تناول مختلف الأنواع تأتي في التحليل الموضوعي لها، فان الصورة التي تعتمد على حاسة النظر تحتوي على ثلاثة مستويات من التحليل الموضوعي، الأول يتمثل في العناصر الأساسية المكونة للصورة التي تقع عليها العين أول ما ترى الصورة، أما الثاني فهو العناصر الثانوية المكملة للصورة وتكن في خلفية الصورة، أما المستوى الثالث فهو الخاص بالمعنى والمفاهيم التي تحويها الصورة و التي تشترك كل العناصر في التعبير عنها. وبذلك فان الصوت والفيديو ينتظران من يتبنى البحث بهما في مواصفات الشكل والمحتوى.

ومن النماذج نظم استرجاع الصور:

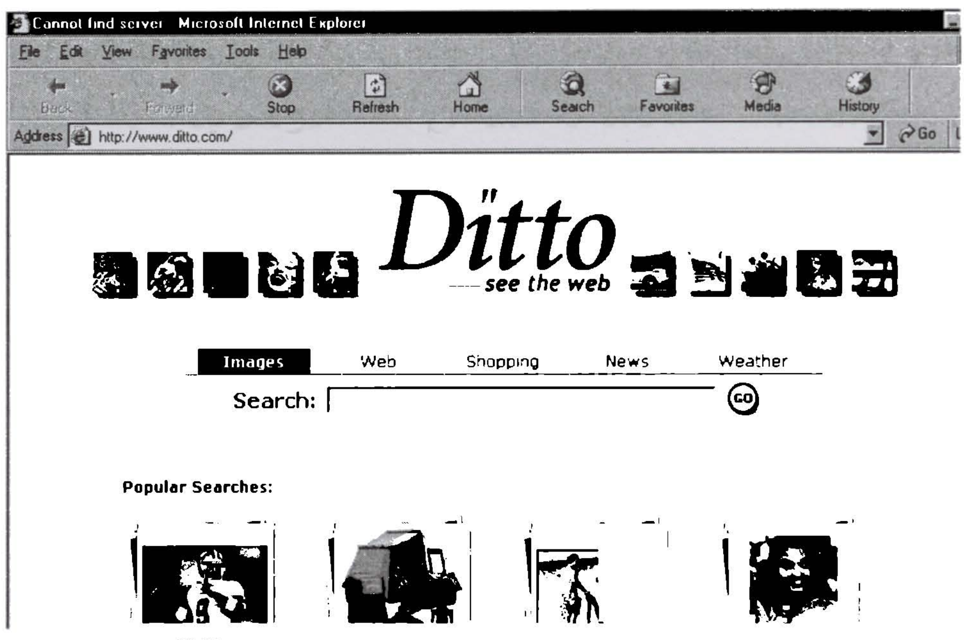
(٢) يتناول الباحث في الدراسة ثلاثة محاور يضمها عنوان الدراسة " محركات بحث الصور على الإنترنت : دراسة تحليلية لوضع مواصفات محرك بحث صور ثابتة يدعم خصائص اللغة العربية " أما المحوران الآخران فهما مواصفات محركات البحث ، و الآخر هو خصائص استرجاع الصور في اللغة العربية .

نظام الاسترجاع : Ditto

يعد www.ditto.com "Ditto" من أهم محركات بحث الصور على الإنترنت، وهو من المحركات المتخصصة في شكل الصور. يبحث المحرك في أنواع من الصور مثل : الصور الفوتوغرافية والرسومات واللوحات الفنية وغيرها من العناصر المكملة لمصطلح Image . يستخدم Ditto برنامج الزاحف في تكوين مجموعاته بالإضافة إلى قاعدة البيانات ، حيث يحدد برنامج الزاحف محتوى الصفحات حديثة الالتقاط وهو الإجراء الذي يعتبر أول مراحل تحليل محتوى صفحات الصور، والتعرف إلى مجموعة الكلمات التي يمكن استرجاع الصور بها . هذا إلى جانب الاعتماد على تحديد مدى الصلة لمجموعة الصور موضع الاستدعاء، مما يمكن من شأنه تحقيق درجة عالية من الدقة في بحث المستفيد . إن ما سبق يعني استخدام المحرك لمجموعة الكلمات التابعة للصورة في التعرف إلى محتوى الصورة الموضوعي . ويتيح محرك البحث التصفح الموضوعي العام عن الاسترجاع، إضافة إلى إمكانية البحث بالكلمات المفتاحية، وإن كان الباحث يعتقد أنها أسلوب الاسترجاع الرئيس لصور قاعدة البيانات . أما عرض النتائج، فإنه يعتمد على شكلين أساسيين هما عرض الصور بجانب البيانات التابعة لها، والآخر هو عرض وحدات الصور التي تضاهي كلمات البحث كلقطات thumbnails يمكن للتعرف إلى الصور ومن ثم يمكن الانتقال إلى الشكل الأصلي للصورة[1].

---

(١) Ditto : Privacy Policy , Ditto , 2002 , cited 10/10/2002 , http://www.ditto.com/privacy.asp



الشكل رقم (٥ –٥) واجهة استخدام محرك البحث (Ditto).

نظام الاسترجاع : Corbis

يعد موقع البحث www.corbis.com أهم مواقع بحث الصور على العنكبوتية، يرجع ذلك إلى عوامل كثيرة من أهمها حجم المجموعات التي تخضع للبحث على الخط المباشر ، وتقدر هذه المجموعة بـ ٢.١مليون صورة من حجم كلي يصل إلى ٦٥ مليون صورة. ويؤدي هذا الموقع عمله في بحث الصور من خلال تجميع كم كبير سواء على الخط المباشر أو قواعد البيانات المكتنزة ، يهدف أكثر ما يهدف إلى الربح. لم يشر الموقع إلى إجراءات العمل إلا أنه لا يعد محرك بحث كما يعكس المفهوم من طبيعة أداء محركات البحث ، جاء ذلك بالنظر إلى طريقة تجميع المواد المصورة التي تعتمد في معظمها على الإضافة اليدوية إلى قاعدة البيانات . من خلال مجموعة البيانات التي ترافق عرض الصورة في التسجيلة النهائية المكونة من المادة المصورة ،

جملة واصفة للصورة في شكل عنوان، ثم المصور، ورقم الصورة؛ إلا أن هذا المحرك الاعتباري يسمح باستخدام أسلوبين من البحث هما : البحث بالتقسيم الموضوعي أو التصفح Browsing لرؤوس الموضوعات، ولا يتبع هذا التقسيم أي خطط تصنيف محددة إلا أنه يعبر عن الموضوعات بشكل عام . أما الأسلوب الآخر فهو المتمثل في البحث بالكلمات المفتاحية، مع إمكانية استخدام آليات بحث مثل ؛ استخدام علامات التنصيص " " لتطابق المصطلحات كما كانت في شكل البحث ، إضافة إلى الاستخدام التلقائي لمعاملات البحث البوليني AND , OR . كل ما سبق يعمل على الوصول إلى درجة عالية من الدقة المطلوبة في البحث.

٥ / ١ / ٢ / ١ / ٣ محركات بحث الصوت: Sound Search Engines

تعد ملفات الصوت MP3 هي أشهر وسيط للتعامل مع الموسيقى على العنكبوتية ، وبالإضافة إلى ترميز الصوت فان هذه الملفات تتسم أيضًا بوجود مدى محدود من البيانات، هذه الملفات تسمح من خلاله بإضافة نص أو كلمات وحقول مصاحبة لملفات الصوت Meta data ، وتوجد لمثل هذه الملفات الكثير من الأشكال Formats يحتوي على حقول لوصف المادة الصوتية باستخدام الكلمات المفتاحية منها :

أولاً: ID3v1

إن ملفات الـ MP3 في شكلها الأصلي لا تحتوي على مساحة لكتابة حقول الوصف، إلا أن الباحث Eric Kemp قد حرص على إرفاق مثل هذه الحقول مع المادة الصوتية في مساحة بسيطة ليخرج الشكل ID3 . تجمع هذه الحقول مختلف البيانات الببليوجرافية عن المادة الصوتية مثل عنوان الأغنية أو

المادة الصوتية، اسم المطرب أومن يحل محله ،الألبوم ، السنة " تاريخ النشر "، التعليق، ومساحة ١ كيلو بايت لكتابة نوع المادة ؛ موسيقى كلاسيك ، روك Rock ، وطنية ... الخ . وما لبث أن تطور المعيار على يد Michael Mutschler إلى شكله ID3v1 بإضافة معلومات عن رقم مسار الصوت على القرص المليزر؛ إلا أن الباحث قد لاحظ في هذه الحقول غياب الحقل أو الترميز الخاص بالموضوع ووصف محتوى المادة المسموعة .

ثانيًا : ID3v2 :

قامت مجموعة من الأشخاص على رأسهم Martin Nilsson في عام ١٩٩٨م، بوضع مجموعة من الحقول في أشكال تناسب كل حقل على حدة، ومن خلالها توضع البيانات الببليوجرافية الخاصة بالمادة المسموعة هي ID3v2 ، إلا أنها توضع قبل الصوت في الملف الحاوي لها بعكس ما كان عليه ID3.(١)

ومن نظم استرجاع الصوت على الويب:

نظام الاسترجاع : Find Sounds

يعتبر أحد محركات البحث التي تختص بالبحث عن المواد الصوتية، يحاكي محرك "www.findsounds.com" مختلف الاستخدامات لكل أعمار مستخدمي العنكبوتية . يعتمد محرك البحث على برنامج الزاحف كطبيعة محركات العنكبوتية في اقتناء صفحات العنكبوتية ، إلا أنه يعمل بشكل مخصص على ملفات المواد الصوتية ، فضلا عن اختيار أنواع

---

(١) Report on MP3 Searching . Search Tools consulting, 2002, cited 2/11/2002 , http://www.searchtools.com/info/mp3-search.html.

محددة من الملفات الصوتية مثل AIFF, AU, and WAVE التي تستخدم مع أرضيات تشغيل مثل Macintosh, Unix, and Windows، والتي يشيع استخدامها على العنكبوتية . يعتمد هذا المحرك أيضًا على الكلمات التابعة للمادة الصوتية في ملفات الصوت على العنكبوتية لاستخدامها في عملية التكشيف ومن ثم استدعاء الصوت بحسب كلمات البحث المدخلة من جانب المستخدم . أما عرض النتائج، فإنه يتمثل في عرض العشر وحدات الأولى من مجموع النتائج التي طابقت كلمات البحث، إلى جانب ذلك فإن واجهة النتائج تتضمن خيارات الانتقال إلى صفحات النتائج الأخرى أو فتح ملف الصوت، عرض مسار ملف الصوت على العنكبوتية Web URL إلى جانب حجم المادة الصوتية والتي لا تزيد لهذا المحرك على M 2 كما جاء في خيارات البحث(١).

٥ / ١ / ٢ / ١ / ٤ محركات بحث المواد الحيوية Animation Search Engines:

ارتبط ظهور المواد الحيوية باستخدام الحاسبات الآلية في التعليم والبحث ؛ حيث لا يوجد لمثل هذه المواد أصول تقليدية مثل بقية الوسائط الأخرى. ولا تلقى مواد الرسوم الحيوية Animation نفس المقدار من الاهتمام الذي تناله الأنواع الأخرى من محركات بحث العنكبوتية ، فلا تتيح محركات البحث إمكانية البحث المتخصص في ملفات الرسوم الحيوية ، إلا أقل القليل الذي يبحث داخل هذا النوع من الوسائط . ويرجع الباحث ذلك إلى قلة تواجد هذا النوع من الملفات على العنكبوتية ذلك لأن محركات البحث تعمل مع مختلف أنواع الملفات التي توجد على العنكبوتية بنفس المعايير والإجراءات، مما يعني

---

(١) Find sounds : Search the Web for Sounds , Comparisonics Corporation , 2002 , cited 10/10/2002 , http://www.findsounds.com/help1.html.

أن قصور بحث هذا النوع يرجع إلى تواجد ملفات الرسوم الحيوية في الشكل الرقمي في نطاق البيئة العنكبوتية .

ومن نماذج نظم استرجاع الصوت:

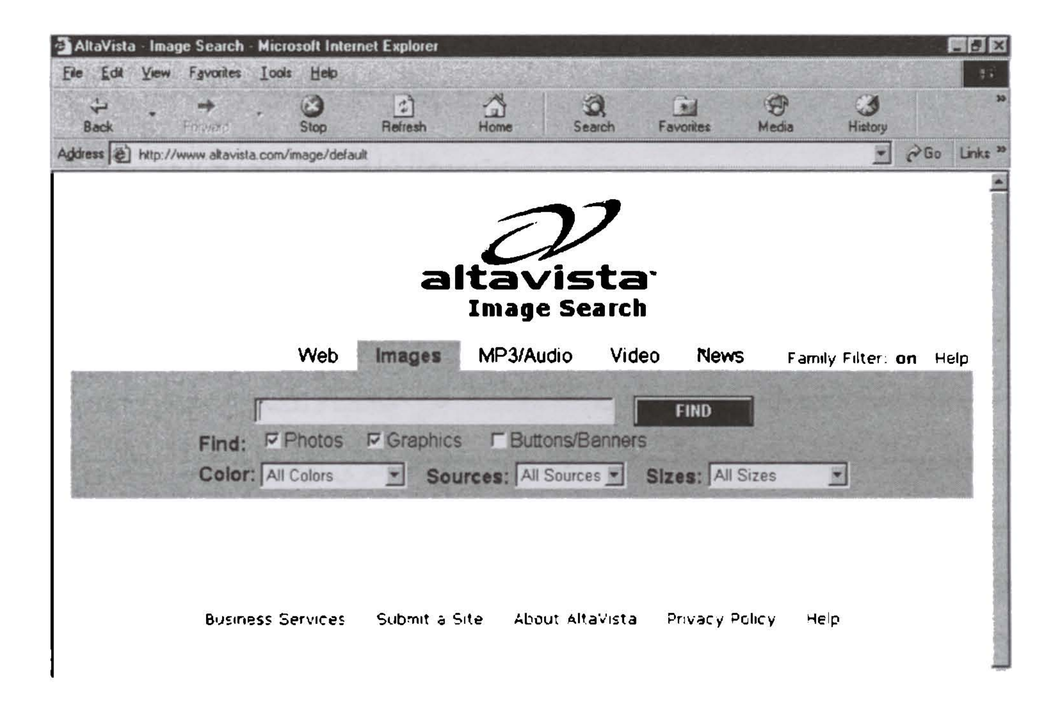
٥ / ١ / ٢ / ٢ نظم عامة لاسترجاع الوسائط المتعددة:

نظام الاسترجاع : Lycos

يشتمل محرك "www.lycos.com" على مجموعة كبيرة من ملفات الوسائط المتعددة، إلا أنه لا يعتمد على آليات بحث متقدمة كتلك التي توجد في غيره من محركات البحث الأخرى، يسير هذا المحرك وفق طبيعة عمل محركات البحث العامة بالاعتماد على عناصر عمله الثلاثة، ألا وهي برنامج الزاحف Crawler أو سفير التعارف لمواقع وصفحات العنكبوتية المختلفة، ثم التحليل الأولى لأنواع تلك الصفحات التي تصنفها إلى صفحات نصية تخضع للبحث العام أو تلك التي تندرج أسفل أحد أنواع ملفات الوسائط المتعددة. ثم تكشف تلك الصفحات بالتعرف إلى أهم الكلمات الدالة على محتوى المادة سواء المرئية أم المسموعة، يلي ذلك دور آليات البحث في استدعاء تلك الملفات بحسب الكلمات المفتاحية الداخلة في إستراتيجية البحث .

نظام الاسترجاع : AltaVista

يعتمد محرك AltaVista كما سبق على آليات بحث متقدمة تختلف بطبيعة نوع وسط المعلومات موضع البحث. على سبيل المثال؛ فانه في حالة بحث ملفات الصوت يقدم آليات اختيار نوع الملف AIFF أو AU أو WAVE ، إلى جانب مدة تشغيل ملف الصوت . أما في حالة بحث الصور Images فانه يقدم آليات مثل لون الصورة ، نوع الصورة مثل الصور الفوتوغرافية أو رسومات .



الشكل رقم (٥ –٦)واجهة استخدام محرك البحث (**Altavista**).

نظام الاسترجاع : Fast

تتبع أداة البحث http://www.multimedia.alltheweb.co " Fast " محرك البحث المتعدد all the web ؛ حيث يعد Fast الوسيط المسئول عن بحث ملفات الوسائط المتعددة من صور وفيديو وصوت . ولا تختلف هذه الأداة في بحث ملفات الوسائط عن مثيلاتها السابقة من محركات البحث في العمل سواء كان التجميع لهذا النوع من الملفات أو التكشيف، إلا أن الاختلاف قد يرجع فقط إلى تغير بعض آليات البحث، وهو الأمر الذي تتسم به محركات البحث عامة.

٥ / ١ / ٣ نظم استرجاع نظم الويب:

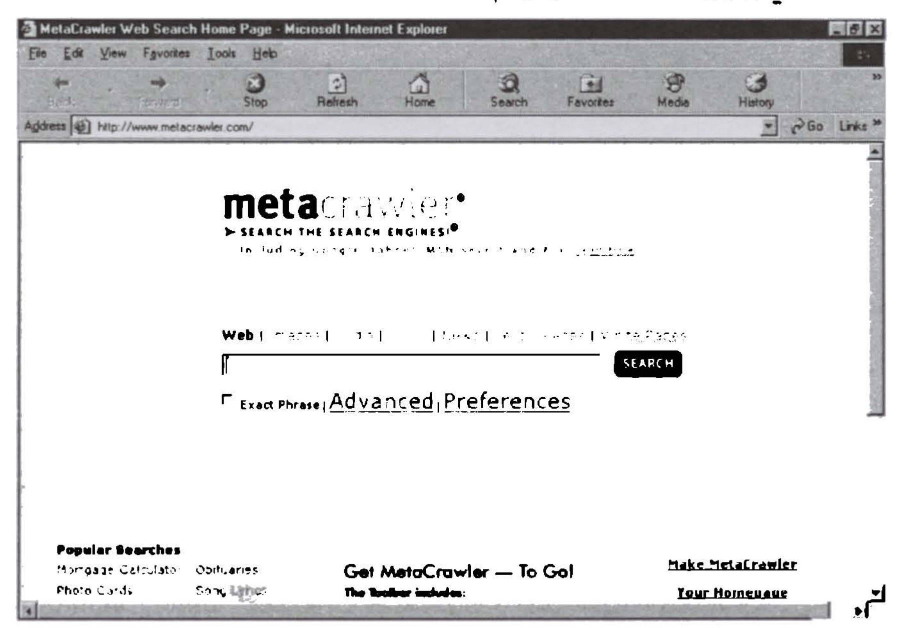
يقوم هذا النوع من نظم استرجاع الويب على معالجة واسترجاع محتوى نظم الاسترجاع الأخرى التي تتعامل مع صفحات وملفات الويب، ومن ثم فهي تقدم لمستخدم الويب محتوى أكثر من قاعدة بيانات يمكن للمستفيد البحث داخلها واسترجاع معلوماتها في وقت واحد، خلاف البحث في كل نظام على حدة. وللوصول إلى هذه الغاية فإن نظم استرجاع نظم الويب تعمل على توفير واجهة بحث عامة تستقبل استفسار البحث وتوجيهه إلى مختلف النظم المرتبطة بها كمحركات البحث أو قواعد البيانات. وتنقسم هذه النظم إلى نوعين بحسب طبيعة نظم الاسترجاع التي تتعامل معها.

٥ / ١ / ٣ / ١ نظم استرجاع محركات البحث " محركات البحث المتعددة":

يعرف قاموس المكتبات والمعلومات على الخط المباشر ODLIS البحث المتعدد على أنه " بحث للمعلومات باستخدام برامج صممت لبث واسترجاع المعلومات من مجموعة محركات بحث مختلفة ثم دمج وإعادة ترتيب النتائج المسترجعة. ومن أهم أمثلة محركات البحث المتعددة على الويب Dogpile, Monster Crawler, and ProFusion". أيضا عرفت شركة Blue Angel Technologies عملية البحث المتعدد على أنها " عملية بحث المعلومات اعتمادًا على طرح سؤال بحث واحد لبحث مجموعة من نظم المعلومات المختلفة في وقت واحد. ومن الممكن أن تشمل هذه النظم فهارس المكتبات وقواعد البيانات التجارية والمجموعات الرقمية والمجموعات الخاصة". وقد ركز أحد التعريفات الأخرى على محرك البحث ذاته دون عملية البحث كما سبق في المفهومين السابقين ؛ حيث عرف محرك البحث المتعدد على أنه" أحد أنواع أدوات البحث التي تمكن مستخدم الويب من بحث مجموعة من اثنين أو

أكثر من محركات وأدلة البحث في وقت واحد. وهذه الأداة لا تملك كشافًا لمصادر المعلومات خاصًّا بها ، وإنما صممت لتكون على قمة أدوات بحث الويب الأخرى". ويرى الباحث أن هذا التعريف الأخير يعكس بدقة ماهية محرك البحث المتعدد(1)(2)(3).

ومن أهم نماذج محركات البحث المتعددة نموذجا all the web و meta crawler الذي يعبر عنه الشكل رقم (٥ –١٤):



الشكل رقم (٥ –٧) واجهة استخدام نظام استرجاع محركات البحث(Metacrawler).

(١) Reitz, Joan M. metasearch, ODLIS — Online Dictionary for Library and Information Science, Libraries Unlimited, 2006, cited 20/9/2007, cited at http://lu.com/odlis/odlis_m.cfm# metasearch.

(٢) metasearch.Blue Angel Technologies, Inc.2005, cited 25/9/2007, cited at http://www.BlueAngelTech.com.

(٣) metasearch.Blue Angel Technologies, Inc.2005, cited 25/9/2007, cited at http://www.BlueAngelTech.com.

يوضح الشكل رقم (٥ –١٤) نموذجًا لمحركات البحث المتعددة كنظم لاسترجاع المعلومات، وهي تتسم بمجموعة من الخصائص مثل: -

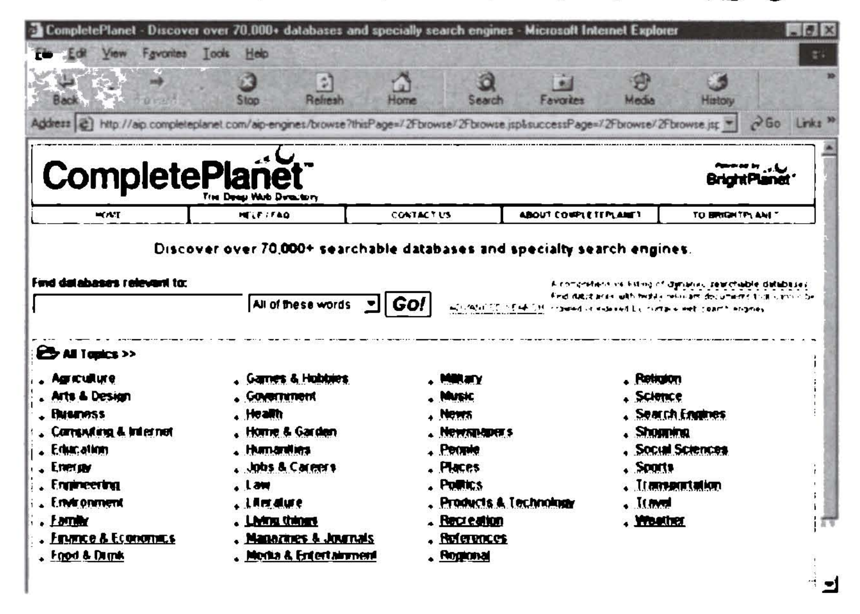
- لا تضم محركات البحث المتعددة مجموعات النظم الأخرى داخل مستودعاتها، وإنما تعمل على الربط بينها وبين مستخدمي الويب.
- تتسم نظم محركات البحث المتعددة بتوفير واجهة بحث واحدة تتعامل مع مختلف آليات البحث في النظم الأخرى.
- تتميز هذه النظم عن محركات بحث الويب العادية بتوفير قدر أكبر من الشمولية والتغطية لمعلومات موضوع البحث.
- تعمل هذه النظم على إعادة صياغة وترتيب النتائج المسترجعة من نظم المحركات الأخرى، لعرضها على واجهة واحدة أمام المستفيد.

٥ / ١ / ٣ / ٢ نظم استرجاع قواعد البيانات:

تختص هذه النظم بتنظيم واسترجاع محتوى قواعد بيانات الويب، التي يطلق عليها الويب غير المرئية أو العميقة invisible web, deep web. وقد جاءت هذه النظم لمعالجة القصور في نظم الويب المعتمدة على برامج الزاحف، التي لا تستطيع الوصول إلى محتوى قواعد الويب والتعامل مع ملفاتها، هذا إلى جانب حجم المعلومات الهائل الذي تحويه قواعد البيانات ؛ حيث زادت أعداد قواعد البيانات على الويب إلى الدرجة التي أصبح حجم هذه القواعد يفوق حجم صفحات وملفات الويب الأخرى بما يقدر بخمسمائة مرة. ومن ثم شكلت قواعد البيانات نوعًا جديدًا من مصادر الويب يمثل عائقًا أمام محركات بحث الويب، يحتاج بطبيعته وخصائصه إلى نوعٍ متطور من نظم استرجاع مصادر

الـويـب، وقـد تمـثل هـذا الـنوع ـ في محـركات وأدلـة بحـث الـويـب غـير المـرئية invisble web search engines and directories.

وتتشابه نظـم اسـترجاع قـواعد بـيانات الـويب ـ في عملـها مـع نظـم محـركات الـبحث المـتعددة ؛ حيـث تـسترجع الأولى المحـتوى الداخـلي لقـواعد البـيانات دون أن تخـصص لـه مـستودعًا خاصًـا لحـفظ هـذه المعلـومات، وهـو مـا يـتم ـ في محـركات الـبحث المـتعددة مـن اسـترجاع محـتوى محـركات الـبحث الأخـرى وإعادة ترتيبها ـ في تـرتيب عـام أمـام المستفيد ـ في أثناء عـرض النتائج. ومن أهم أمثلة نظم استرجاع قواعد البيانات على الويب invisible-web, infomine, turbo10, completeplanet.



الشكل رقم (٥ –٨) واجهة استخدام نظام استرجاع محركات البحث(completeplanet).

٥ / ٢ الخاتمة:

تتعدد وتتنوع نظم استرجاع معلومات الويب بحسب طبيعة الوثائق التي تحويها أو مستوى مصادر المعلومات الذي تتعامل معه، وقد وضح من خلال هذا الفصل

مدى التنوع والتباين في العمل بين نظم استرجاع الويب. ومن ثم فإن هذا الاختلاف بين نظم استرجاع الويب وبعضها البعض، يعد أكبر تأكيد على مدى الاختلاف في طبيعة المعالجة والعمل وآليات الاسترجاع بين ما تكون عليه نظم استرجاع الويب ونظم الاسترجاع الأخرى التقليدية أو على الخط المباشر. وقد عكس هذا الفصل أيضًا مدى التطور الذي تشهده نظم استرجاع الويب من حيث ظهور نوع جديد منها يمكنه البحث والاسترجاع من داخل نظم استرجاع أخرى، وهو ما كان متمثلا في نظم محركات البحث المتعددة ونظم استرجاع قواعد بيانات الويب غير المرئية.

الفصل السادس

قياس كفاية نظم استرجاع المعلومات

٦ / ٠ التمهيد:

تسعى مختلف أنماط نظم استرجاع المعلومات التقليدية أو الرقمية، على الويب أو على الخط المباشر إلى تحقيق أفضل درجة كفاية في العمل على معالجة المعلومات وتقديمها إلى المستفيدين. ولكي تستمر نظم استرجاع المعلومات في المحافظة على كفايتها أو التعرف إلى نقاط الضعف والقوة في عملها، فهي تلجأ دائما على استخدام مبدأ التقييم والتقويم؛ حيث يكون التقييم للتعرف إلى خصائص عمل النظام وقياس درجات الأداء في المعالجة والاسترجاع، أما التقويم فهو عمليات التغيير والتعديل التي يقوم بها النظام للتحديث والتطوير والتحسين من أداء العمليات المختلفة داخله. ويهتم هذا الفصل بالتعرف إلى المقاييس والمعايير المستخدمة لقياس الأداء لنظم استرجاع المعلومات، سواء تلك التي تخص النظام أو البيئة أو المستفيد وحصوله على المعلومات. وقد وضحت الفصول السابقة أن التغذية المرتدة جزء لا يتجزأ من عوامل بناء نظام الاسترجاع، وتتمثل التغذية المرتدة في التعبير عن رضاء أو تعليق المستفيد لإيضاح درجات الرضا أو السخط تجاه المعلومات التي حصل عليها في شكل نتائج مسترجعة من النظام.

لقد تحدثت موضي بنت إبراهيم الدبيان عن أهمية القياس والتقييم لخدمات المكتبات، وبيان العوامل المؤثرة في تقييم وقياس كفاية الأداء في المكتبات. والحديث هنا عن المكتبات إنما هو حديث عن نوع من أنواع نظم استرجاع المعلومات في شكلها التقليدي، أي أنه يتعلق في بداية الأمر ونهايته بتقييم أداء نظم الاسترجاع. أشارت إلى أن القياس والتقييم يهدفان إلى تبيان مواطن الضعف والقوة في خدمات المكتبات ومراكز المعلومات حتى يصل

التقييم إلى درجة تمكن نظام استرجاع المعلومات من رفع كفاية الخدمة وتطويرها كماً وكيفاً، فالتطوير يبدأ بالقياس والتقييم وينتهي بالتقويم[1].

وهناك عوامل كثيرة تؤثر بالقياس من أهمها:

١- طبيعة الأشياء المراد قياسها: فهناك أشياء يمكن قياسها بطريقة مباشرة في المكتبات ومراكز المعلومات مثل خدمات الإعارة الخارجية، وهناك أشياء من الصعوبة قياسها مثل كمية الإعارة الداخلية.

٢- كفاية القائمين على القياس: ومدى إلمامهم بأساليب القياس ومناهجه المختلفة ودقتهم في جمع البيانات اللازمة لعمليات التحليل الإحصائي ووضوح الرؤية أمامهم بالنسبة لأهداف عملية القياس.

٣- أساليب القياس المستخدمة : هل هي أساليب كمية أم نوعية. فالأساليب الكمية أكثر دقة ووضوحاً؛ لأنها تعتمد على الحقائق والأرقام مثل الإحصاءات المكتبية، أما الأساليب النوعية فتخضع للعوامل الشخصية.

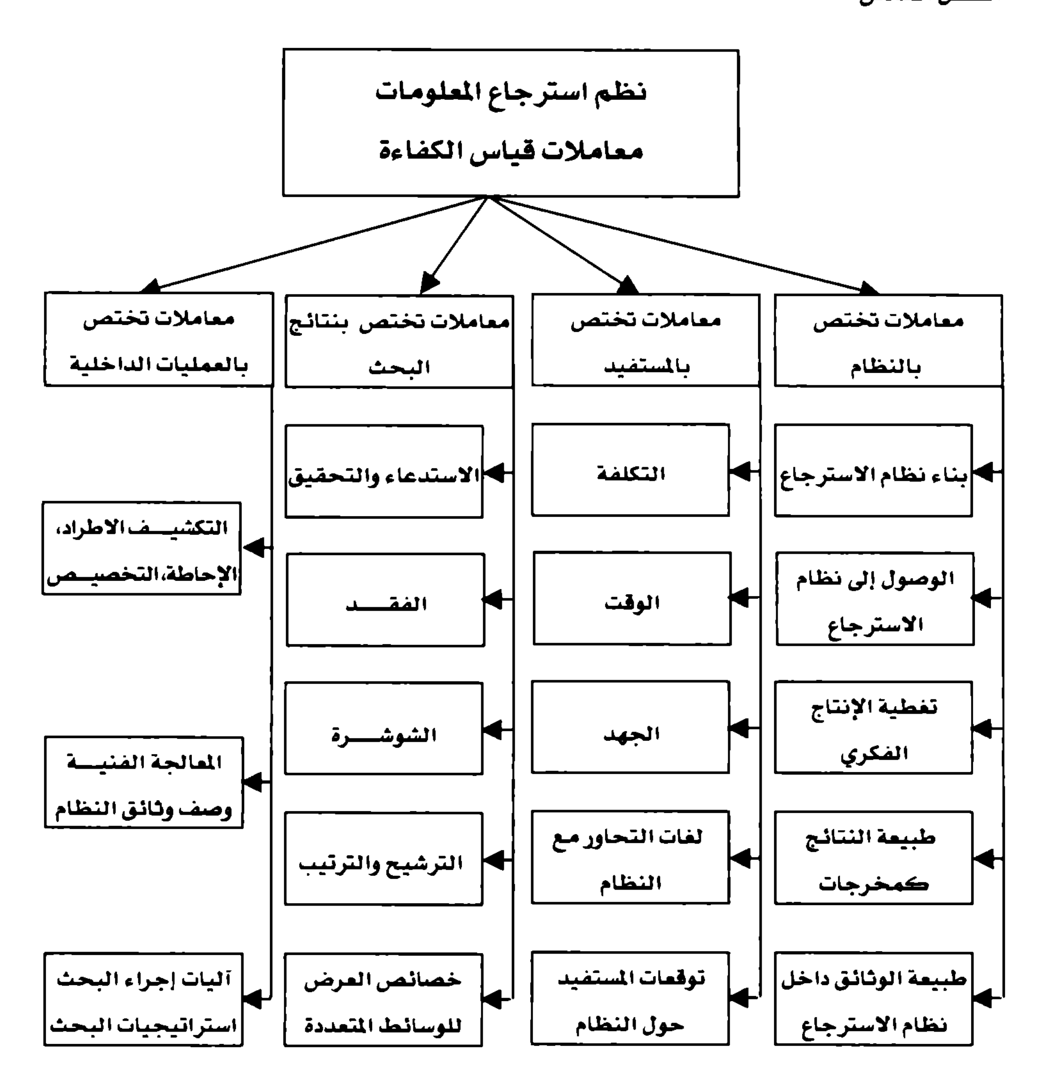
٦ / ١ مُعَاملات قياس الكفاية لنظم استرجاع المعلومات

تحتوي نظم استرجاع المعلومات على عناصر فرعية متشابكة تقوم بأداء مهام ووظائف النظام، بالإضافة إلى احتواء النظام على مجموعة من العمليات

---

(١) موضي بنت إبراهيم الدبيان. قياسات أداء خدمات المكتبات ومعايير تقييمه، **مجلة المعلوماتية**، تاريخ المقالة تاريخ الاستشهاد ٢٠٠٧/١١/١١ ، مسار http://informatics.gov.sa/modules.php?name=Sections&op=printpage&artid=122

التي تتكامل لخدمة المستفيدين، وعلى ذلك يمكن القول إنه لقياس كفاية نظم استرجاع المعلومات لا يمكن أبدًا الاعتماد على منظور واحد أو النظر من جانب واحد لنظام استرجاع المعلومات. وانطلاقًا من أن كل عنصر داخل النظام قد يساعد على نجاح أو فشل الغاية التي يريد النظام إدراكها، فإنه يجب تحديد مقياس لأداء عمل ونجاح كل عنصر من عناصر العمل داخل النظام. ولذلك فقد عمد الباحث هنا إلى الحديث عن مجموعة من معاملات قياس الكفاية، تقسم بحسب مختلف العناصر العمليات التي تتشارك جميعًا في صنع وبناء نظام استرجاع المعلومات. ووزعت هذه المعاملات تبعًا لأربع مجموعات أساسية هي معاملات تختص بنظام الاسترجاع ذاته، وأخرى تخص العمليات والأنشطة القائمة، وثالثة تخص المستفيد وما يرتبط به، وأخيرة تتعلق بطبيعة النتائج المسترجعة. وحري بالذكر هنا أن هذه النظرة الشاملة المتعلقة بقياس كفاية نظم الاسترجاع تعبر عن وجهة نظر الباحث، لما يراه من وجوب الاهتمام بمختلف عناصر بناء النظام دون التركيز على معاملات التحقيق والاستدعاء. بالإضافة إلى ذلك فإن الباحث لم يصادف أيًا من الدراسات التي تحدثت عن قياس الكفاية وقد تناولت نظم الاسترجاع من هذا المنظور الشامل.وقد جاءت المجموعات جميعها ممثلة في الشكل رقم (٦-١):



الشكل رقم (٦ -١) التقسيم العام لمعايير تقييم نظم استرجاع المعلومات

أثر التباين بين نظم استرجاع المعلومات على الويب وغيرها من النظم الأخرى، في تزايد اهتمام الباحثين بمعاملات لقياس كفاية وأداء لمحركات بحث الويب كأهم أنواع نظم استرجاع المعلومات في فضاء الويب. وقد وضعا Gwizdka, Jacek and Chignell, Mark بعض جوانب الحكم على أداء وكفاية

محركات بحث الويب كنظم لاسترجاع المعلومات في بيئة الويب، ومن بين هذه الجوانب[1]:

١- قياس مدى التغطية التي تتسم بها قواعد بيانات محركات البحث؛ إضافة إلى سرعة التحديث والإضافة بها.

٢- قياس إمكانات البحث داخل المحركات، ويحدد ذلك بطبيعة آليات البحث الموجودة داخل نظام الاسترجاع، مثل آليات المنطق البولياني وغيرها.

٣- قياس كفاية الأداء اعتمادًا على استخدام معاملات القياس الكمية للنتائج، مثل معاملات التحقيق والاستدعاء ووقت البحث.

٤- قياس كفاية عرض النتائج المسترجعة، مثل أعداد النتائج في كل واجهة، والبيانات المصاحبة، والخدمات المتاحة للتعامل مع نتائج البحث المسترجعة.

٥- قياس الجهد والوقت المبذولين من المستفيد لقاء التفاعل مع نظام الاسترجاع والحصول على النتائج الملائمة لتلبية الحاجات الموضوعية.

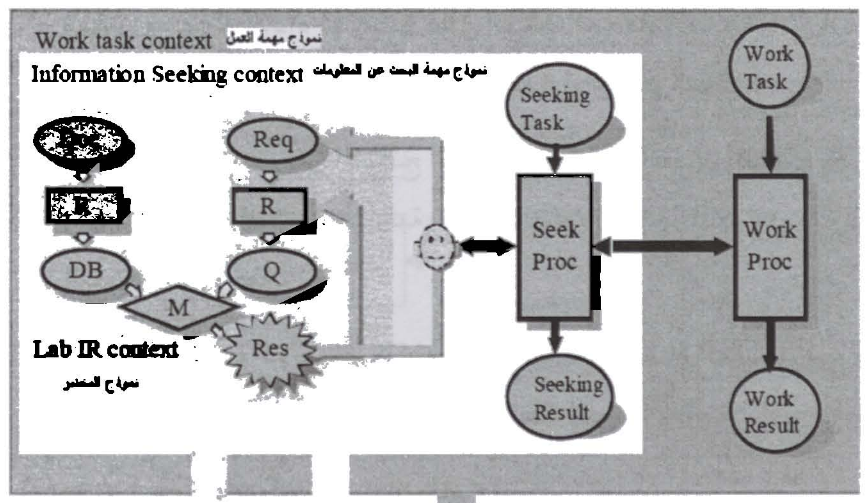
٦ / ١ / ١ مُعَاملات تختص بالنظام :

تتركز معاملات قياس كفاية نظام الاسترجاع في نظرة شاملة وعامة لنظام استرجاع المعلومات، سواء بالجوانب التي تتعلق ببنية النظام عامة، أو بيئة العمل التي يوجد بها نظام الاسترجاع كالبيئة الرقمية أو التقليدية. ومن ثم فإن مختلف المعاملات التالية تتعلق بالكيان العام لنظام استرجاع المعلومات.

---

(١) Gwizdka, Jacek and Chignell, Mark. Towards Information Retrieval Measures for Evaluation of Web Search Engines, University of Toronto, 1998, cuted 15/11/2007, cited at http://anarch.ie.utoronto.ca/~jacekg/pubs/webIR_eval1_99.pdf

وقد انتبه بعض الباحثين مثل Jaana Kekinen & Kalervo Jrvelin لأهمية تقييم نظم استرجاع المعلومات بنظرة شاملة تربط بين التفاعل بين النظام والمستفيد، وبين ما يدور بذهن المستفيد وطبيعة الاحتياجات الموضوعية التي أتى بها إلى النظام؛ إضافة إلى تقييم العمليات الداخلية لنظام استرجاع المعلومات. وهو ما أطلق عليه تقييم المعمل The Laboratory Model. ويوضح الشكل رقم (٦-٢) أن تقييم نظام استرجاع المعلومات يعتمد أطرًا مختلفة يتعلق الإطار العام منها بعلاقة النظام والمستفيد، ويتعلق الإطار الأضيق بالتفاعل الداخلي للعمليات في نظام استرجاع المعلومات[١].



الشكل رقم(٦ - ٢) الأطر المختلفة لتقييم نظم استرجاع المعلومات

---

(١) Kekinen, Jaana & Jrvelin, Kalervo. EVALUATING INFORMATION RETRIEVAL SYSTEMS UNDER THE CHALLENGES OF INTERACTION AND MULTIDIMENSIONAL DYNAMIC RELEVANCE, University of Tampere, 2003, cited 1/11/2007, cited at

يوضح الشكل السابق أن تقييم نظم استرجاع المعلومات بما له من علاقة باحتياجات المستفيدين يجب النظر إليه في ثلاثة أطر أساسية وهي كالتالي:

الإطار الأول : بيئة عمل المستفيد ومنبع احتياجاته ؛ حيث تمثل الاحتياجات للمعلومات الدافع الرئيس وراء استخدام المستفيدين لنظم استرجاع المعلومات، وكما هو معلوم مسبقًا أن هذه الحاجات قد تكون مرتبطة بغرض بحثي أو عملي مرتبط بعمل المستفيد ، أو الحصول على المعلومات لأغراض المعرفة والثقة العامة. ومن ثم فإن مدى تلبية النظام لهذه المعلومات وتغطية هذه الدوافع جميعًا هو الإطار العام للحكم على نجاح أو فشل نظم استرجاع المعلومات في تلبية احتياجات المستفيدين. فضلا عن إمكانية تقييم النتائج والمعلومات كمخرجات من النظام عند تطبيقها واستخدامها في أغراضها المستخدمة. ويدلل على ذلك استخدام النموذج لمصطلحات مثل : work task, work processing, work result.

الإطار الثاني : استخدام النظام ؛ يتمثل الإطار الثاني لتقييم نظم استرجاع المعلومات في التوافق بينه وبين المستفيدين عند الاستخدام وإجراء عمليات البحث المختلفة. فبعد أن استقر المستفيد على النظام الملائم، تأتي مرحلة التفاعل مع نظام الاسترجاع طرحاً للأسئلة وتلقياً للإجابات والنتائج المسترجعة. ومن ثم ينعكس ذلك كله على إثر استخدام المعلومات في مواضعها وقيمة النتائج المسترجعة في تغطية الاحتياجات الموضوعية. ويعبر النموذج عن هذا الإطار ومعناه باستخدام مصطلحات مثل الاستخدام والجودة في usability, qualaty of information, versatility.

الإطار الثالث : أداء النظام ؛ ويتمثل إطار تقييم النظام في التعرف إلى جوانب الضعف والقوة من تحليل العمليات الفنية الداخلية لنظام استرجاع المعلومات، ومن العمليات الأهم تحليلا في النظام هي عمليات المعالجة والتنظيم، وآليات وإستراتيجيات البحث المستخدمة. حيث تبدأ هذه العمليات بتلقي الاستفسارات البحثية وتحليلها وإجراء عمليات المضاهاة على الكلمات المفتاحية، على الجانب الآخر يتم تلقي الوثائق المضافة ومعالجتها وتنظيمها لبناء قاعدة بيانات النظام. وينعكس أثر مختلف هذه العمليات جليًّا على النتائج المسترجعة ومدى نجاح نظام استرجاع المعلومات في استرجاع الملائم من المعلومات واستبعاد غير المتصل بالموضوع. وقد استخدم النموذج السابق للدلالة على هذا الإطار مصطلحات التحقيق والاستدعاء precision, recall, efficiency.

ولم ينه الباحثان دراستهما قبل أن يضعا بعض النقاط التي تتعلق بدراسة وتقييم نظم استرجاع المعلومات، مثل :

١- الحاجة للوقوف على الرابط المشترك بين عمليات تحديد منزلة الوثائق ومدى صلتها بموضوع البحث، والشكل الذي يكون عليه ترتيب النتائج في واجهات عرض نتائج البحث.

٢- الحاجة للتعرف إلى خصائص واجهات بحث النظام على اعتبار أنها ممثل نظام استرجاع المعلومات في التحاور مع المستفيدين، وتأثير ملامحها المختلفة على نجاح التفاعل بين المستفيد والنظام.

٣- الحاجة لتحديد أهداف واضحة عند بناء نظام استرجاع المعلومات، وما القدرات المتوقعة في تلبية احتياجات المستفيدين من المعلومات.

٤- العمل على تحسين معايير التقييم لنظم استرجاع المعلومات، دون النظر فقط إلى خصائص محددة للنظام. فضلا عن تطوير هذه المعايير نظرًا للتغيرات التي تطرأ على نظم استرجاع المعلومات يومًا بعد آخر.

وتمثل العناصر التالية ملامح عامة لتقييم نظم استرجاع المعلومات يمكن الاعتداد بها قبل أو في أثناء قيام النظام بمهامه، وهي:

٦ / ١ / ١ / ١ بناء النظام:

إن بناء نظام استرجاع المعلومات دائمًا يسعى لتحقيق هدف بث المعلومات للمستفيدين بحسب احتياجاتهم الموضوعية. وعلى ذلك فإن الأهداف التي يبنى لأجلها نظام هي ذاتها عوامل التقييم الأساسية ؛ حيث يأتي نجاح البناء أو فشله من تحقيق الأهداف أو عدم الوصول إليها. ولكي تعكس بنية النظام الأهداف المرجو تحقيقها، فإن ثمة عوامل يجب مراعاتها في أثناء التخطيط لبناء نظام الاسترجاع منها:

- ملاءمة نظام استرجاع المعلومات من حيث الوثائق والشكل الرقمي أو التقليدي للبيئة التي يعمل فيها نظام الاسترجاع.
- توفير مختلف الأجهزة والموارد المادية أو البشرية الأخرى التي يحتاجها النظام لأداء مهامه في معالجة وتنظيم وبث المعلومات.
- العمل دائمًا على إصباغ نظم استرجاع المعلومات المراد بناؤها، بمختلف التقنيات التي انتهى إليها مجال نظم الاسترجاع.
- مراعاة المستويات المختلفة من المستفيدين المتوقع تعاملهم مع النظام.

- تحديد الجوانب الأمثل لبناء واجهات البحث، في حالة تعامل نظام الاسترجاع مع شكل أو أكثر من الوسائط المتعددة.

٦ / ١ / ١ / ٢ الوصول إلى النظام:

تسعى نظم استرجاع المعلومات دائمًا للوصول إلى المستفيدين وإتاحة أكبر قدر من المعلومات بأقل جهد يبذله المستفيدون. ويختلف الوصول إلى نظام استرجاع المعلومات بحسب طبيعة وبيئة العمل، فإذا ما كان نظام الاسترجاع يعمل في بيئة تقليدية مثل مختلف مرافق المعلومات، يصعب على المستفيدين الوصول إليه. أما في حالة التعامل مع نظم الاسترجاع سهلة التنقل مثل نظم الاسترجاع على الأقراص المضغوطة CDs، فقد يسهل على المستفيد الحصول عليها والتفاعل معها دون الالتزام بوقت أو مكان. وتقدم بيئة الويب نمطًا مختلفًا للوصول إلى نظم استرجاع المعلومات ؛ حيث تتواجد نظم استرجاع الويب في الوقت والمكان الذي يمكن للمستفيد فيه الدخول إلى فضاء الويب، أي أن نظام الاسترجاع هنا هو الذي يصل إلى المستفيد حيث كان. ويمثل عامل الوصول إلى نظم الاسترجاع أحد معايير المستفيد في حكمه على نظام الاسترجاع بالشكل الإيجابي أو السلبي.

٦ / ١ / ١ / ٣ التغطية للإنتاج الفكري:

تتجه معظم نظم استرجاع المعلومات الآن إلى التخصص في المجال الموضوعي، وقد أتى ذلك نتيجة لتزايد أعداد نظم استرجاع المعلومات على مستويات مختلفة، خاصة في البيئة الرقمية، فعلى مستوى التخصص النوعي لمصادر المعلومات ؛ قد أوجدت الويب نظمًا لاسترجاع المواقع والصفحات

النصية، إلى جانب التعامل مع أنواع الوسائط المتعددة الأخرى من ملفات الصوت والصورة والصور المتحركة. أما على المستوى الموضوعي فقد أضحت محركات البحث المتخصصة تغطي مختلف موضوعات المعرفة البشرية. هذا إلى جانب ظهور نوع آخر من التغطية في مجال نظم استرجاع المعلومات وهو التغطية لمستويات مصادر المعلومات ؛ فهناك من نظم الاسترجاع على الويب ما يهتم بتنظيم ومعالجة صفحات الويب، ونظم أخرى تهتم بالمواقع ذات الصفحات المتعددة، وثالثة تهتم بتنظيم معالجة قواعد البيانات، وأخيرة تهتم بتنظيم محتوى محركات البحث ذاتها كمحركات البحث المتعددة.

تحتاج نظم استرجاع المعلومات دائمًا إلى التحديث من وثائقها لكي تتماشى مع تطور المجال الموضوعي المخصص. فيمثل وقوف نظام استرجاع المعلومات عند مرحلة زمنية محددة في المجال الموضوعي، بمثابة توقف لعمل نظام استرجاع المعلومات ككل. ومن النماذج الأكثر وضوحًا على ذلك نظم استرجاع المعلومات على الأقراص الضوئية أو المدمجة optical or compact discs ؛ حيث لا تقبل هذه النظم تحديثًا أو تطويرًا لوثائقها، مما يصرف المستفيدين عن استخدامها في وقت معين. وهذا بالإضافة إلى تغطية الإنتاج الفكري للمجال المخصص بكافة أشكاله وفي مختلف فروعه الموضوعية. ويؤثر معيار التغطية للإنتاج الفكري في احتياج بعض المستفيدين دائمًا إلى نسبة عالية من الاستدعاء للوثائق دون الاهتمام بعامل التحقيق ومدى اتصال النتائج المسترجعة بموضوع البحث المحدد. ومن ثم يكون معيار التغطية للإنتاج الفكري واحدًا من العوامل المؤثرة في تقييم نظم استرجاع المعلومات.

تتسم محركات بحث الويب كنظم لاسترجاع المعلومات بالمحاولة النظرية لتغطية الإنتاج الفكري الصادر في المجال الموضوعي المخصص أو على المستوى العام، وذلك اعتمادًا على طبيعة محرك البحث. فمن الناحية النظرية تستخدم محركات البحث برامج الزاحف أو العنكبوت لإضافة ما يستحدث من معلومات في المجال الموضوعي، فضلاً عن المتابعة الدائمة للتحديثات التي تطرأ على وثائق الويب. غير أنه من الناحية الواقعية فإن التزايد المستمر لحجم الوثائق على الويب كمًا ونوعًا يجعل من الصعب جدًا على محركات البحث الوصول إلى مرحلة تغطية كل الإنتاج الفكري في بيئة الويب. ويعد معيار التغطية للإنتاج الفكري عاملاً مميزًا عند تقييم نظم استرجاع المعلومات على الويب، عنه في نظم استرجاع المعلومات الأخرى.

٦ / ١ / ١ / ٤ مخرجات النظام:

يشير مصطلح مخرجات النظام هنا إلى جانبين أساسيين هما ؛ **الأول :** طبيعة الوثائق التي يضمها النظام والتي تخرج في شكل نتائج مسترجعة عند إجراء عمليات البحث. فتحتوي نظم استرجاع المعلومات على مستويات كثيرة من المعلومات تتدرج من التعامل مع التسجيلات الببليوجرافية إلى المستخلصات، ثم انتهاءً بالنصوص الكاملة. وعلى ذلك فإن النتائج النهائية من المعلومات تؤثر في إقبال المستفيد أو إعراضه عن استخدام نظام استرجاع المعلومات. ويمكن القول إن الويب قد أحدثت تغيرًا كبيرًا في قدرات المعالجة والتخزين لنظم استرجاع المعلومات، بل والتحول من التعامل مع مستويات الإشارات الببليوجرافية إلى التعامل مع النصوص الكاملة. فلا يمكن القول الآن إن أحدًا من المستفيدين يقبل باستخدام نظام استرجاع معلومات يتقدم به

فقط إلى نصف احتياجاته من المعلومات، وهو تقديم حصر بالإشارات الببليوجرافية للإنتاج الفكري الملائم لموضوع البحث. بل يسعى المستفيد دائمًا لاستخدام نظام استرجاع المعلومات الذي يقدم المعلومات الأولية التي تلبي الاحتياجات الموضوعية.

أما **الجانب الثاني** : فهو يتعلق بالشكل المادي لمخرجات النظام ؛ فقد تعمل بعض النظم على تقديم النتائج والمعلومات النهائية للمستفيدين في دقة تامة، غير أن نظام استرجاع المعلومات لا يتيح للمستفيد استخدام خدمات مثل الطباعة أو التحميل أو النسخ... وغيرها من العمليات التي تيسر للمستفيد الاستخدام الأمثل لمخرجات النظام من المعلومات. ومن ثم تكون المخرجات معيارًا مهمًّا من معايير تقييم وقياس كفاية عمل نظم استرجاع المعلومات.

٦ / ١ / ١ / ٥ طبيعة الوثائق داخل النظام:

تختلف أنواع ومستويات الوثائق التي تختزنها نظم الاسترجاع، ومن ثم تختلف الملامح التي تعمل بها نظم الاسترجاع بحسب طبيعة كل نوع أو مستوى ؛ حيث لا يمكن أن تتساوى خصائص بحث وعرض التسجيلات الببليوجرافية مع النصوص الكاملة، كما لا يمكن بناء ملامح عمل واحدة تتوافق مع خصائص النصوص والوسائط المتعددة داخل نظام استرجاع واحد. وعلى ذلك كانت الملامح التي يعمل بها نظام الاسترجاع نابعة من طبيعة وسيط المعلومات ذاته. ويشكل ذلك تأثيرًا في الحكم على كفاية نظام الاسترجاع عندما يخلو نظام استرجاع للمادة المصورة من آليات التنظيم أو البحث الملائمة للصور، مما ينعكس بدوره على حكم المستفيد ببعض القصور في أداء نظام الاسترجاع.

لقد ذكر لانكستر Lancaster أن هناك بعضًا من الأخطاء التي قد تشوب عمل نظام استرجاع المعلومات، بما ينعكس سلبًا في نهاية الأمر على رضاء المستفيدين عن النتائج المسترجعة من نظام المعلومات. ومنها[1]:

- عدم تخصص مصطلحات المكنز المستخدم بالدرجة الكافية.
- وجود أخطاء في قاعدة البيانات ومن ذلك الأخطاء الإملائية أو اللغوية .
- تأخير تغطية التسجيلات الببليوجرافية للوثائق.
- عدم كفاية التغطية.
- إخفاق القائمين بعمليات تحليل المحتوى والتنظيم في أثناء عمليات التكشيف من النواحي التالية:

❖ فشل المكشف في تحليل المفهوم بطريقة صحيحة وعدم فهمه موضوعها بدقة.

❖ فشل المكشف في ترجمة المفاهيم واختياره مصطلحات غير صحيحة للتعبير عن الموضوعات.

❖ حذف المكشف لجوانب مهمة من المفاهيم التي تنطوي عليها الوثيقة.

❖ استخدام المكشف مصطلحات عامة للتعبير عن الوثيقة بسبب الافتقار إلى المصطلحات المتخصصة الملائمة.

---

(١) Lancaster , F. W. Vocabulary Control For Information Retrieval .- 2nd ed .- Arlington, Virginia : Information Resources Press , 1986 .- p148.

❖ استخدام المكشف مصطلحات عامة للتعبير عن المفاهيم على الرغم من توافر المصطلحات المتخصصة.

٦ / ١ / ٢ مُعَامِلات تختص بالمستفيد:

يعد المستفيد أهم عوامل بناء نظام استرجاع المعلومات، بل يمكن القول إنه لا يمكن لنظم استرجاع المعلومات العمل أو الاستمرار دون التفاعل مع المستفيدين. ومن ثم فإن الدور الذي يقوم به المستفيد في التفاعل مع النظام لا يقتصر فقط على استرجاع المعلومات، وإنما المشاركة في تقييم وتطوير نظم استرجاع المعلومات التي يستخدمها. ويتمثل دور المستفيد في تحديد بعض النقاط المؤثرة في قدرات النظام على تلبية احتياجاته من المعلومات، وهذه النقاط قد يعبر عنها قولاً أو فعلاً. وتسمى جميعها بالتغذية المرتدة feedback. فمن غير الممكن لنظام الاسترجاع العمل من جانب واحد هو معالجة واسترجاع المعلومات، دون النظر إلى مدى نجاحه في القيام بذلك وهو ما لا يمكن التعرف إليه إلا من خلال المستفيدين أنفسهم.

لا يختلف الأمر في التقييم من خلال المستفيد بين نظم استرجاع المعلومات التقليدية والرقمية، فقد أشار Bernard J. Jansen في دراسة له عن تقييم أداء محركات البحث إلى أهمية الاستعانة بآراء المستفيدين، وذلك لما لها من قلة التكلفة عن بعض قياسات الأداء الأخرى، فضلاً عن استقبال النظام لها في حال إجراء عمليات البحث. ويتم ذلك من خلال تتبع العمليات المختلفة التي يقوم بها المستفيد من نسخ وطبع واختيار بين النتائج. وقد أشار الكاتب أيضًا إلى وجود بعض البرامج التي تعمل باتجاه المستفيد client side يمكنها حساب

كم النتائج المختارة ومن ثم تحديد نسبة التحقيق والاستدعاء في نتائج النظام المسترجعة أمام المستفيد. ويقسم المستفيد النتائج المسترجعة إلى ثلاث مجموعات بحسب أهميتها وارتباطها بموضوع البحث هي[1] :

١- مجموعة النتائج المرتبطة بالموضوع.

٢- مجموعة النتائج غير المرتبطة بالموضوع.

٣- مجموعة نتائج أخرى. وقد تزيد عليها مجموعة رابعة هي:

٤- مجموعة النتائج التي تتصل بالموضوع وتم الوصول إليها.

سلوك المستفيد ونظم استرجاع المعلومات الويب :

يختلف سلوك المستفيد في التعامل بين نظم استرجاع الويب Web IR users والنظم الأخرى، وذلك لتحقيق التفاعل في أقصى حالاته بين المستفيد والنظام عند استخدام محركات البحث. تتطلب محركات البحث كنموذج لنظم الويب إلى اعتماد المستفيد على مهاراته البحثية وكفايته في التعامل مع واجهات البحث دون اللجوء إلى الوسيط في أي من مراحل البحث. ومن ثم فإن هناك مجموعة من العوامل المؤثرة في تحقيق نجاح عمليات البحث داخل نظم الويب، فالبيئة الديناميكية، والتحديث السريع المتلاحق، وآليات المنطق

(١) Jansen, Bernard J. & Sharma, Himanshu. Automated Evaluation of Search Engine Performance via Implicit User Feedback, The Pennsylvania State University, 2004, cited 15/11/2007, cited
http://ist.psu.edu/faculty_pages/jjansen/academic/pubs/jansen_automated_evaluation.pdf

البولياني، والبحث داخل النصوص الكاملة، والصياغة الذاتية لاستراتيجيات البحث، تعد جميعًا مؤثرات في التفاعل بين المستفيد والنظام.

ومن ثم عملت الكثير من الدراسات مثل Kirsch and Silverstein باستخدام القياسات الرياضية لمجموعات الاستفسارات البحثية المطروحة على محركات بحث مثل Infoseek, Excite, Altavista على دراسة وتقييم سلوك مستخدمي محركات البحث، والتفاعل بين المستفيد والنظام[1].

إن ثمة وجهتي نظر يجب الانتباه إليهما في قياس كفاية الأداء من جانب المستفيد ؛ الوجهة الأولى تخص المستفيد الذي يهتم بالتكلفة التي أنفقها للحصول على المعلومات، أو الوقت الذي استغرقته عمليات البحث والوصول إلى المعلومات، أو الجهد المبذول لقاء الحصول على النتائج المرضية. أما وجهة النظر الثانية فتتمثل في نظام الاسترجاع الذي يهتم بآراء المستفيد للتعرف إلى جوانب الضعف والقوة في كفاية المعالجة والاسترجاع للمعلومات بما يلبي حاجات المستفيد بدقة ويسر. وسوف يتناول الباحث هنا وجهة النظر الأولى من خلال مجموعة من العناصر تتمثل في التالي:

٦ / ١ / ٢ / ١ التكلفة:

قد يحصل المستفيد على المعلومات التي يرغبها من نظام الاسترجاع، بل وقد تتسم هذه المعلومات بكل الملامح الإيجابية في المعلومات من حداثة وشمولية وتغطية وتعمق في موضوع الباحث، وفقًا لما يرغبه المستفيد، غير أن المستفيد في نهاية الأمر قد لا يصل إلى درجة الرضاء المرجوة، وهذا لا يعود إلى

---

(١) Jansen, Bernard J. LBID.

طبيعة المعلومات المسترجعة، وإنما إلى التكلفة التي تكبدها المستفيد في حصوله على المعلومات. وعامل التكلفة قد يؤثر بدرجة كبيرة على النظر إلى كفاية النظام ؛ حيث قد يتطلب الخروج بالمعلومات الدقيقة تكلفة أعلى، والعكس من ذلك في قبول المستفيد لنتائج متواضعة نظير انخفاض عامل التكلفة. ومما لا شك فيه أنه على المستفيد التعرف إلى خصائص النظام والصعوبات التي قد يواجهها في استرجاع المعلومات، فهناك من النظم ما هو متاح على الأقراص الضوئية الأقل تكلفة غير أنها قد تحتوي على معلومات راجعة أو غير محدثة، وفي المقابل التعامل مع النظم ذات التكلفة المرتفعة للوصول إلى المعلومات الحديثة.

٦ / ١ / ٢ / ٢ الوقت response time

يمثل الوقت الجانب الأكثر أهمية لدى المستفيد عند الحاجة إلى المعلومات بطريقة سريعة. وقد يكون الدافع وراء السرعة في الحصول على المعلومات هي مرحلة البحث التي يمر بها المستفيد، أو قد تكون الحاجة للانتهاء من كتابة تقرير أو مشروع. ومن ثم يمثل الوقت عاملاً لنجاح أو فشل المستفيد في الحصول على المعلومات، وتبعًا لذلك يكون حكم المستفيد على نظام الاسترجاع. ويشكل الوقت قاعدة أساسية عند كثير من المستفيدين تقوم على أن كفاية نظم استرجاع المعلومات لا تتمثل فقط في إخراج النتائج المتصلة والمطابقة لموضوع البحث، وإنما الحصول على هذه النتائج بأسرع وقت ممكن.

ويتضمن نظام استرجاع المعلومات الكثير من العوامل التي من شأنها إطالة أو تقليل الوقت اللازم للخروج بالمعلومات. يحصر منها الباحث ما يلي:

١- إن البيئة التي يوجد بها النظام تؤثر بشكل كبير في الوقت اللازم لإجراء عمليات البحث ؛ حيث يقاس الوقت اللازم لإجراء عمليات البحث بالثواني المعدودات في حالة التعامل مع نظم استرجاع الويب، خلاف النظم التقليدية أو الإلكترونية الأخرى فيمتد من ساعات إلى أيام.

٢- يشكل التفاعل المباشر بين المستفيد والنظام عاملاً إيجابيًا في تقليل الوقت اللازم للخروج بالمعلومات، أما التفاعل غير المباشر المعتمد على الوسيط فمن شأنه إطالة وقت التفاعل ؛ نظرًا للمرور بمراحل مثل الاتفاق على الكلمات البحثية، وإجراء تجارب لاستراتيجيات بحث مختلفة، ثم استعراض النتائج وقبولها أو رفضها.

٣- تشكل ملامح المستفيد ذاته عاملاً مؤثرًا في وقت التفاعل مع نظم استرجاع المعلومات ؛ فقد تكون المهارات البحثية وإتقان آليات البحث ذات المستوى المنخفض هو عامل إطالة لوقت البحث والتفاعل مع النظام.

٦ / ١ / ٢ / ٣ الجهد:

يقطع المستفيدون رحلة الوصول إلى المعلومات بداية من الشعور بالحاجة إلى المعلومات، وانتهاءً بالحصول على المعلومات الملائمة لتلبية البحث. غير أن هذه الرحلة قد تأخذ من المستفيد قدرًا كبيرًا من الجهد المبذول للحصول على ما يريده من معلومات. ويختلف القدر المبذول من جهد المستفيد باختلاف نظام استرجاع المعلومات المنتقى لاسترجاع المعلومات ؛ فعندما يملك المستفيد القدرة

على صياغة الكلمات البحثية الدالة على الموضوع مع امتلاك المهارات البحثية وفهم آليات البحث، عندئذ يمكن للمستفيد الوصول إلى المعلومات المرجوة بأقل جهد ممكن، خلاف ما يمكن أن يكون الأمر عليه مع الأنواع الأخرى من نظم الاسترجاع أو في حالة الاعتماد على وسيط لإجراء البحث. ويمكن تصور الجهد المبذول من المستفيد للحصول على المعلومات في المواضع التالية:

- تحتاج النظم التقليدية التي تبعد جغرافيًا عن المستفيد جهدًا كبيرًا للوصول والتفاعل مع النظام، ناهيك عن استخدام الوسيط لإجراء البحث. ويقل الجهد في حالة نظم استرجاع المعلومات على الخط المباشر، وأقل منه في حالة نظم استرجاع الويب.

- تمثل لغات المعالجة والتنظيم والبحث المعقدة التي تعتمدها بعض نظم استرجاع المعلومات، عاملا لزيادة جهد المستفيد في التحاور مع النظام والتفاعل معه، بل وجهدًا أكبر في حالة انخفاض مستوى إدراك لغات البحث وإستراتيجياته.

- يعتمد التفاعل غير المباشر بين المستفيد ونظام استرجاع المعلومات على استخدام الوسيط، وقد يمثل ذلك عاملاً لبذل الجهد من جانب المستفيد لشرح وتوضيح الحاجات الموضوعية لديه، ثم إجراء عمليات البحث والتطلع في النتائج المسترجعة للتعرف إلى الملائم منها.

٦ / ١ / ٢ / ٤ التحاور مع النظام / الوسيط:

يختلف المستفيدون فيما بينهم في مستوى التعرف إلى طبيعة الحاجات المعلوماتية عند الوصول إلى نظام الاسترجاع، ويؤثر هذا الاختلاف على درجة

نجاح التحاور والتفاعل بين المستفيد والنظام. ومن ثم فإن استيعاب المستفيد لموضوع البحث من شأنه تحقيق النجاح أو الفشل في الحصول على المعلومات الملائمة. ويبدأ نجاح التفاعل مع نظام الاسترجاع عند تحديد الكلمات الدالة وصياغة إستراتيجية البحث. ويتم إدخال كلمات البحث المفتاحية إلى النظام اعتمادًا على أسلوبين أساسيين هما:

❑ التحاور مع الوسيط:

يبلغ دور الوسيط من الأهمية الدرجة التي قد تسبب نجاح أو فشل وصول المستفيدين إلى ما يرغبونه من معلومات، حيث يكون الوسيط منوطًا بالتعبير عن الحاجات الموضوعية المحددة المشكلة في ذهن المستفيد، فضلاً عن كونه شارحًا للمستفيد إمكانات النظام وخصائص المعلومات المتوقع الحصول عليها من نظام الاسترجاع. ويقوم المستفيد بالتحاور مع الوسيط عند فقده القدرة على استخدام النظام مباشرة وإتقان لغة بحث النظام. وحري بالذكر هنا أن التفاوت في الملامح الإدراكية للموضوع، والصفات الشخصية بين المستفيد والوسيط قد يسير بالمستفيد إلى الحصول على معلومات أو نتائج غير متصلة بموضوع البحث.

❑ التحاور مع النظام:

يتمثل التحاور مع النظام في وجود لغة تحاور مشتركة بين المستفيد والنظام، تمكن المستفيد من التحاور المباشر دون استخدام الوسيط. والإلمام بلغات التحاور مع النظام يعتمد في الفهم والإدراك لمجموعة آليات البحث وإستراتيجية الربط بين الكلمات المفتاحية في أثناء عملية البحث والاسترجاع.

وتعد آليات المنطق البولياني ومعاملات البحث الحر wild card, bolean logic أهم أنواع لغات البحث المستخدمة مع نظم استرجاع المعلومات.

٦ / ١ / ٢ / ٥ التوقعات حول النظام :

لا يحمل المستفيد حاجاته من المعلومات فقط عندما يذهب لنظام الاسترجاع، وإنما يحمل معه أيضًا توقعات كثيرة حول النتائج التي سيمنحها له نظام استرجاع المعلومات. وعلى إثر هذه التوقعات، فإن المستفيد يصبغ كل تفاعله مع النظام بتقديراته المسبقة، ويبدأ ذلك في محاولة التعبير عن موضوعه بالشكل العام أو المخصص عند التحاور مع الوسيط، أو اختيار كلمات بحث محددة قد تبعد أو تقترب من موضوع البحث في مرحلة صياغة إستراتيجيات البحث. وإذا ما أردنا أن نفند طبيعة توقعات المستفيد حول النظام فإنها تتعلق بالآتي:

- توقعات المستفيد حول مدى تغطية الإنتاج الفكري داخل نظام الاسترجاع، فإذا ما كانت التغطية واسعة فعليه استخدام مصطلحات بحث أضيق للاهتمام بمستوى التحقيق وتفادي الاستدعاء. والعكس من ذلك في استخدام كلمات ذات دلالة واسعة لزيادة الاستدعاء في حالة تغطية النظام المحدودة للإنتاج الفكري.
- توقعات المستفيد حول كفاية الوسيط ككونه حلقة وصل بين المستفيد والنظام، وقدرته على استيعاب الأفكار التي يحملها المستفيد، وقدرته أيضًا على التعبير عن هذه الأفكار بالكلمات الدالة المناسبة.

- توقعات المستفيد حول طبيعة المعلومات التي سيخرج بها من نظام الاسترجاع، فهل تقف هذه المعلومات عند حد الإشارات الببليوجرافية أم تتعدى ذلك إلى المعلومات الأولية النهائية.

إن مختلف العوامل السابقة المتعلقة بالمستفيد هي أهم معايير التقييم والكفاية من وجهة نظر المستفيدين ؛ ذلك لكونها ترتبط بآمال المستفيد في الحصول على المعلومات بأقل قدر من الجهد، وأقل تكلفة ،وفي أسرع وقت إذا ما دعت الحاجة لذلك، هذا إلى جانب الحصول على المعلومات الأكثر اتصالاً بالموضوع. ويمكن القول : إن عدم تحقق واحد من العوامل السابقة، قد يكون بمثابة حكم المستفيد على أداء نظام الاسترجاع بالفشل في معالجة واسترجاع المعلومات. ومن ثم فإنه على مصممي نظم الاسترجاع وضع العوامل السابقة كأهم معاملات تقييم وقياس كفاية نظام استرجاع المعلومات.وبقي هنا أن نضع أمثلة للأخطاء التي قد يقوم بها المستفيد نفسه عند التفاعل مع نظام استرجاع المعلومات والتي من شأنها أن تؤثر على قيمة النتائج المسترجعة. ومنها :

عدم تحديد المستفيد للحاجات المعلوماتية الحقيقية خلال مقابلة ما قبل البحث.

- عدم جودة بناء إستراتيجية البحث.
- الإخفاق في إجراء المقابلة المرجعية من قبل المسئول.
- البحث تحت مصطلحات خاطئة.
- اختيار قاعدة بيانات غير ملائمة.

٦ / ١ / ٢ مُعَامِلات تختص بالعمليات:

تهتم مختلف دراسات التقييم وقياس الكفاية لنظم استرجاع المعلومات بمعاملات محددة تتركز في التحقيق والاستدعاء؛ إضافة إلى الوقت و الجهد. وهذه المعاملات ترتبط بالنتائج المسترجعة في نهاية عمليات البحث، مما يعني أننا نهمل العوامل التي ترتب عليها خروج هذه النتائج بالشكل الذي خرجت عليه ؛ أي أن زيادة نسبة الاستدعاء في نتائج البحث قد ترتبت على اتباع منهج خاص في عمليات المعالجة والتنظيم لوثائق النظام، ومثالا على ذلك فإننا عند اتباع منهج الشمولية والإحاطة في عملية تكشيف الوثائق داخل النظام، فإن هذا يعني التعبير عن مختلف الموضوعات الرئيسة والفرعية الثانوية داخل الوثيقة، والتعبير عنها بمصطلحات كشفية، وعند بناء إستراتيجية البحث واستخدام أي من الكلمات الكشفية الدالة على الموضوعات الثانوية فإنها ستعمل على استدعاء واسترجاع الوثائق محدودة الاتصال بالموضوع، بما يزيد عامل الاستدعاء أمام المستفيد. والعكس من ذلك في اعتماد منهج التخصيص في عملية التكشيف فإنه يرفع نسبة التحقيق في النتائج المسترجعة.

ويرى الباحث أنه للتحكم في نسبتي التحقيق والاستدعاء في النتائج النهائية المسترجعة، فيجب الاهتمام بصياغة منهج محدد عن إجراء عمليات المعالجة والتنظيم لوثائق نظام استرجاع المعلومات. ولكي يتضح ذلك، فإننا نتناول معًا خصائص عمليات المعالجة والتنظيم للوثائق، وبيان أثر ذلك على نسب التحقيق والاستدعاء النهائية عند الاسترجاع. وتتمثل المعالجة والتنظيم في العمليات التالية:

٦ / ١ / ٣ / ١ التكشيف:

تعتبر عملية التكشيف محور عمل المعالجة والتنظيم داخل نظم استرجاع المعلومات، وقد بدأت واستمرت عملية التكشيف في كونها إحدى عمليات تحليل المحتوى لإخراج الواصفات الدالة على موضوع أو موضوعات الوثيقة. غير أن البيئة الرقمية المتمثلة في الويب قد أضافت بعدًا آخر لعملية التكشيف، فلم يعد يقتصر الأمر على البحث بالكلمات الموضوعية، إنما تعدى ذلك للبحث بالواصفات الدالة على الشكل والأبعاد والخصائص المادية الأخرى، وهذا إنما يعود إلى تنشيط مختلف حقول التسجيلة الببليوجرافية وإمكانية استرجاع الوثيقة من النظام بمختلف الحقول الواردة بها. ومن ثم فقد تكون الواصفات الدالة على الشكل هي الحد الفاصل بين استرجاع وثيقة واستبعاد أخرى في نفس موضوع البحث. وتتأثر نتائج البحث الممثلة في معاملي الاستدعاء والتحقيق بالخصائص المتبعة في عملية التكشيف. وذلك كما يلي: -

أولاً : التخصيص :

يحمل مصطلح التخصيص مفهومين يتمثلان في :

أحدهما : أن التخصيص يعني اختيار الموضوع أو الموضوعات الرئيسية داخل الوثيقة ثم التعبير عنها بالمداخل الكشفية، دون الاهتمام بالموضوعات الأخرى الثانوية والتي يهتم بها مبدأ الإحاطة أو الشمول، ومن ثم فإن المفهوم الأول هو المقابل الفعلي لمفهوم الإحاطة داخل التكشيف.

وثانيهما : أن التخصيص هو اختيار المصطلحات الأكثر دقة في التعبير عن موضوع أو موضوعات الوثيقة، دون أن يكون المصطلح الكشفي أعم أو أخص من موضوع الوثيقة الفعلي. ويرتبط هذا الجانب بمدى قدرة المكشف على فهم محتوى الوثيقة واختيار المصطلحات الأكثر ملاءمة له.

إن التخصيص يحدد وجوب تكشيف المحتوى تحت المصطلح الأكثر تخصصاً والذي يغطي الموضوع بأكمله، ويكون ارتباط التخصيص في عملية التكشيف بنوع المصطلحات التي توفرها لغة التكشيف للمكشف، مثلاً : إذا كانت المصطلحات المستعملة في الكشاف مباشرة وتصف الموضوعات وتتطابق معها أو تدل عليها بالتحديد فعندها يكون التكشيف مخصصاً. أما إذا كانت الموضوعات الكشفية لا تطابق الموضوعات بالتحديد، فإن اللغة تكون أقل تخصيصاً. ويؤثر بذلك في رفع كم المعلومات المتصلة بموضوع البحث، مما يعني زيادة إيجابية لمعامل التحقيق أمام المستفيد.

ثانيًا : الإحاطة:

يقصد بالإحاطة في التكشيف قدرة المكشف على إدراك مختلف الموضوعات الرئيسة والفرعية التي تعالجها الوثيقة، ثم التعبير عن هذه الموضوعات بمداخل كشفية تحيط بكل جوانب محتوى الوثيقة. وتعتمد الإحاطة أو الشمول على وجهة النظر القائلة بأن كل موضوعات الوثيقة الرئيسة كانت أم الفرعية يجب التعبير عنها بكلمات مفتاحية، ذلك لما يمكن أن يترتب عليه تحقيق الإفادة للمستفيدين واسترجاع كل المعلومات المتصلة بموضوع البحث من قريب أو بعيد.

تتعامل نظم استرجاع الويب متمثلة في محركات البحث مع تكشيف صفحات ومواقع الويب مستخدمة مبدأ الإحاطة والشمول، ويأتي ذلك نتيجة لتكشيف النص الكامل للوثائق واعتبار أن أيًا من الكلمات الواردة في المتن هي مداخل لاستدعاء الوثائق الواردة بها، دون الاهتمام بدلالة أو عمومية الكلمات الواردة في النص. ومن ثم فإننا جميعًا نلاحظ الأعداد الهائلة وحجم الاستدعاء داخل محركات البحث، وهو ما يختلف تمامًا عنه في استخدام أدلة بحث الويب web directories التي تعتمد على العنصر البشري في عملية التكشيف لاستخراج الكلمات الدالة فقط، بما يعني قلة المداخل الكشفية الخاصة بكل وثيقة، ومن ثم انخفاض معامل الاستدعاء وارتفاع عامل التحقيق بين النتائج المسترجعة.

ثالثًا : الاطراد:

والمقصود به التوحيد والاتساق في القرارات التي يتخذها المكشفون في عملية التكشيف. فإذا كان التخصيص عاملاً نوعياً مهمًا في عملية التحليل الموضوعي، فإن الشمول عامل كمي مهم. أما الاطراد فيتأثر بعدة عوامل منها: لغة التكشيف، ومدى الشمول، وحجم لغة التكشيف ومدى تخصيصها، والخبرة التي يتمتع بها المكشفون، ومؤهلاتهم، وكذلك التخصص الموضوعي للوثيقة، ومدى توافر الأدوات المعاونة لدى المكشفين. ويرتبط بالاطراد جودة التكشيف، فيجب أن تكون مصطلحات التكشيف المخصصة لوثيقة ما وكذلك الخطوات التي تتخذ للحصول على المصطلحات موحدة ولو اختلف الأشخاص القائمون بالتكشيف، وأن تكون مستقرة نسبياً خلال عمر نظام التكشيف المعني.

يستطيع نظام الاسترجاع تحقيق الاطراد بين المكشفين بالاعتماد على مجموعة من الأدوات المستخدمة لإتمام عملية التكشيف مثل : قوائم رؤوس الموضوعات، والمكانز، وقوائم الاستناد. وينعكس أثر ذلك عند استرشاد المستفيد بهذه الأدوات لاختيار المصطلحات البحثية وإتمام استراتيجية البحث، أي أن الاطراد هنا سيشمل المستفيد للبحث بنفس الكلمات المستخدمة في التكشيف، مما يحقق معه أفضل درجات التحقيق أو الاستدعاء وفقًا لعوامل التخصيص أو الإحاطة كما سبق ذكره.

٦ / ١ / ٣ / ٢ المعالجة الفنية:

تستخدم نظم استرجاع المعلومات عمليات المعالجة الفنية؛ لتحقيق الوصف الفني للوثائق داخل النظام، وذلك من أجل وصف خصائص وملامح الوثيقة الشكلية والموضوعية أمام المستفيد، فيما يطلق عليه التمثيل المكثف .. هذا إلى جانب استخدام حقول المعالجة الفنية كعناصر نشطة يمكن للمستفيد استرجاع الوثائق بالبحث داخل هذه الحقول. وكما سبق تناوله فإن المعالجة الفنية تستخدم الميتاداتا أو البيانات الخلفية metadata لوصف وثائق النظام. وتتنوع المعايير التي تضمها الميتاداتا، فمنها ما يخص وسيطًا محددًا لحمل المعلومات مثل معيار VRA 3 لوصف ملفات وسائط المعلومات المصورة، ومنها ما يلائم ملفات الويب عامة مثل معياري MARC and Dublincore .

تعرض نظم استرجاع المعلومات نتائجها وفقًا لمجموعة من مستويات العرض، تختلف هذه المستويات بحسب عدد حقول الوصف المعروضة أمام المستفيد بداية من المستوى الموجز إلى المستوى الفني، الذي تعرض فيه التسجيلة في شكلها الكامل. ويمكن للمستفيد باستخدام حقول الوصف

المعروضة داخل النتائج التعرف أكثر إلى محتوى وماهية كل وثيقة بما يمكنه من تحديد ملاءمتها أو عدم ملاءمتها لموضوع البحث.

٦ / ١ / ٣ / ٣ آليات البحث:

تتكامل آليات البحث مع أدوات الربط بين كلمات البحث؛ لتكوين إستراتيجية بحث نظام الاسترجاع بما يتوافق وحاجات المستفيد من المعلومات. وسواء أكانت أساليب البحث المستخدمة هي البحث بالتصفح أم البحث الحر اعتمادًا على أدوات الربط المختلفة، فإنها تمثل لغة التحاور والتفاعل بين النظام والمستفيد، ومن المؤكد أن إجادة المستفيد لهذه اللغات هو نجاح في التحاور مع النظام، نجاح بدوره في استرجاع المعلومات المتصلة بالموضوع، ومن ثم يصبغ المستفيد حكمه على نظام الاسترجاع بالإيجابية. وعندما يجد المستفيد مشقة في استخدام لغات التحاور وصعوبة فهم آليات البحث، فإنه يعجز عن استرجاع المعلومات الملائمة، ومن ثم كان التقييم سلبيًا على كفاية نظم استرجاع المعلومات.

إن من أهم المصطلحات التي تذكر بذكر محركات البحث هو مصطلح إستراتيجيات البحث. ويمثل هذا المصطلح أهمية كبيرة في فهم محركات البحث؛ ذلك لأنها تمثل أحد الأعمدة الثلاثة التي يقوم عليها محرك البحث؛ الإضافة والتكشيف والبحث. وقد ذكر Ross Tyner أن إستراتيجيات البحث يجب أن تقوم على خمسة عناصر أساسية هي[1]:

(١) Tyner, Ross. Effective Search Strategies, staff screening, 2003, cited 20/11/2007, cited at http://staffscreening.det.wa.edu.au/education/cmis/eval/search/search1.htm.

١- أن يتم صياغة أسئلة البحث الصياغة التي تحدد الهدف من طرح السؤال.

٢- أن تحدد المفاهيم المهمة من كل كلمات السؤال المطروح.

٣- أن تحدد مصطلحات أو كلمات البحث التي تصف تلك المفاهيم.

٤- أن تحدد المترادفات المختلفة التي تعبر عن نفس الموضوع والمفاهيم المطلوبة.

٥- أن ينشأ منطق بحث يمكن استخدامه من جانب الباحثين.

٦ / ١ / ٤ مُعَامِلات تختص بنتائج البحث :

تتحقق غاية نظام استرجاع المعلومات عند استرجاع الملائم من وثائقه الداخلية لاحتياجات المستفيد من المعلومات. ويمكن القول إن النتائج النهائية التي تم استرجاعها هي خير شاهد على كفاية أو فشل نظام الاسترجاع في تقديم المعلومات لمستفيديه. غير أن النجاح في الوصول إلى النتائج المرجوة لم يأت فقط نتيجة لاختيار كلمات البحث المعبرة، أو استخدام جيد لآليات البحث، أو انتقاء أمثل للمعلومات من جانب المستفيد، وإنما يأتي هذا النجاح من تكامل مختلف الجوانب السابق ذكرها في تقييم نظام الاسترجاع ؛ فالنجاح في بناء النظام والقدرة على المعالجة الدقيقة، واختيار الملائم لاحتياجات المستفيد المتوقعة من مناهج تكشيف ولغات بحث، كل ذلك يساهم في الوصول إلى النتائج التي يتمناها المستفيد. ويعكس تقييم النتائج النهائية مُعَامِلا التحقيق والاستدعاء، مع الأخذ في الاعتبار أنها دائما ما

تكون من وجهة نظر المستفيد الذي يملك وحده إدراك حدود موضوع البحث، والمسؤول الأوحد عن اختيار الكلمات الدالة المكونة لاستراتيجيات البحث(١).

إن ثمة مصطلحًا لم تتناوله الدراسات القائمة على معاملي الاستدعاء والتحقيق، في حين أن المستفيد يبني عليه إيجابية أو سلبية نسبتي الاستدعاء والتحقيق في النتائج المسترجعة. وهذا المصطلح هو الصلة أو الارتباط بالموضوع the relevance، ومفهوم الارتباط بالموضوع هو المحرك الخفي للمستفيد في الحكم على نظام استرجاع المعلومات ككل، بل والحكم على النتائج المسترجعة بأنها مغالية في الاستدعاء على حساب التحقيق، أو أن النتائج المسترجعة ذات تحقيق عالٍ وتدل على كفاية نظام الاسترجاع في بث المعلومات. وإذا ما أردنا إثبات ذلك فلننظر إلى معنى الاستدعاء والتحقيق، نرى أنهما يعنيان معًا الوثائق المرتبطة بالموضوعات المسترجعة، أي أن معاملي الاستدعاء والتحقيق يدلان على الوثائق المرتبطة أو المتصلة بالموضوع التي تم استرجاعها أمام المستفيد.

تتبنى نظم استرجاع المعلومات على الويب معاملات الاستدعاء والتحقيق لتقييم أداء النظام، غير أن ذلك لا يتم بمعزل عن معاملات أخرى كما هو الحال في نظم الاسترجاع التقليدية والخط المباشر. وقد تحدث Fidel Cacheda and Angel Via في دراسة لهما عن بعض الجوانب الأخرى التي تهتم بها خادمات نظم استرجاع المعلومات على الويب، ويتم تحليل المعلومات المتعلقة

---

(١) Bellot, Patrice. Evaluation of Document Retrieval Systems, University of Avignon , 2001, cited 20/11/2007, cited at http://www.lia.univ-avignon.fr/fich_art/146-lreceval.pdf.

بهذه الجوانب للتعرف إلى حالة وكفاية الأداء لنظام الاسترجاع. ومن هذه الجوانب ما يلي[1]:

١- حساب التاريخ والوقت الذي أجريت فيه عمليات بحث النظام واسترجاع النتائج.

٢- التعريف باستفسار البحث من حيث شكل إستراتيجية البحث، وكيفية البحث من خلال البحث الحر أو البحث بالتصفح.

٣- الوقت المستنفد لعملية البحث بداية من طرح السؤال حتى تلقي النتائج.

٤- تحديد بالأشكال والصور التي تخللت النتائج المسترجعة.

٥- تحديد الوقت الإضافي المستغرق لتحميل الأشكال والصور المنتقاة من النتائج المسترجعة.

وتخصص نظم استرجاع الويب عناصر أخرى لقياس أداء النظام، تتركز في متابعة ما تكون عليه صفحات عرض النتائج. ويتم دراستها من خلال العناصر التالية:

١) عدد المجالات الموضوعية التي يمكن أن تضمها النتائج المسترجعة في صفحة عرض النتائج.

٢) عدد وحدات الوثائق المسترجعة الخاصة بكل مجال أو قطاع موضوعي.

---

(١) Cacheda, Fidel and Via, Angel. Performance evaluation of Web Information Retrieval Systems and its application to e-Business, University of A Coruña, 2003, cited 25/11/2007, cited at http://www.tic.udc.es/~fidel/docs/publications/e-2002.pdf.

٣) عدد وحدات الوثائق التي تظهر في صفحة عرض النتائج ككل.

٤) التنظيم والترتيب الذي وردت به الوثيقة الأولى أو الوثائق الأُول في بداية النتائج.

ويتم تخزين مختلف هذه المعلومات للاعتماد عليها كإحصاءات تستخدم فيما بعد لتقييم وقياس أداء نظام استرجاع المعلومات على الويب.

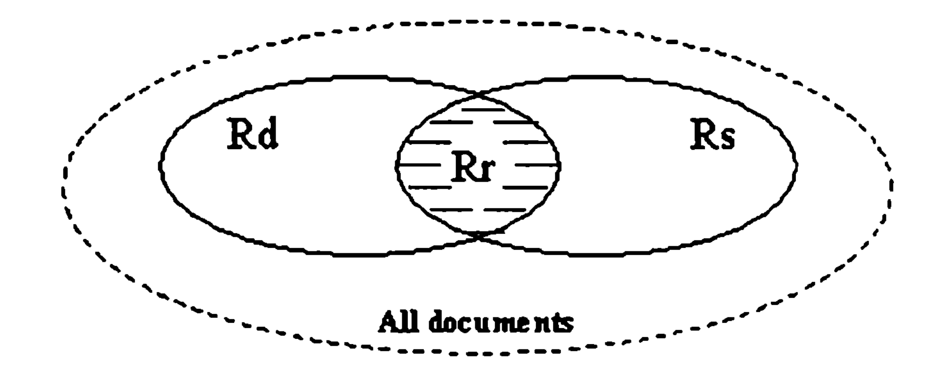
٦ / ١ / ٤ / ١ الاستدعاء والتحقيق : Recall and Precision

لقد آثر الباحث الجمع بين معاملي الاستدعاء والتحقيق عند الحديث عنهما ، وعدم الفصل بينهما كباقي المعاملات الأخرى اللاحقة. وذلك لما رآه الباحث من ارتباط وثيق بين هذين المعاملين ؛ حيث يأخذ كلا المعاملين تعريفاً واحداً ، وهو مجموعة الوثائق المتصلة بالموضوع والمسترجعة من داخل نظام الاسترجاع. أيضًا ، فإن اهتمام المستفيد ينصب بالحالة التي تكون عليها نسب الاستدعاء والتحقيق في النتائج المسترجعة ، كما أن الاستدعاء والتحقيق هما مقياسا الكفاية والنجاح لنظام الاسترجاع من وجهة نظر المستفيد ، فضلا عن كون الاستدعاء والتحقيق هما المعبران عن كل معايير التقييم والقياس الأخرى، فإن صَلُحَا صلحت باقي المعايير، وإن فَشِلا كان الحكم على أداء النظام ككل بالفشل.

ولكي نتعرف إلى معاملي الاستدعاء والتحقيق، فإننا نستعين بالشكل رقم(٦ -٣) ، للتعرف إلى كلا المفهومين والعلاقة بينهما[١]:

---

(١) Nordbotten, Joan C. Multimedia Information Retrieval Systems, 2006, cited at 30/11/2007, cited at
http://nordbotten.com/ADM/ADM_book/ADM-TextRetrieval.htm#R/P.



Rd = relevant documents
Rr = retrieved relevant documents
Rs = response set

Recall = Rr/Rd
Precision = Rr/Rs

الشكل رقم (٦ -٣) يوضح طبيعة معاملي الاستدعاء والتحقيق والعلاقة بينهما.

يوضح الشكل رقم (٦ -٣) طبيعة الاستدعاء والتحقيق من خلال النظرة الشاملة لنظام استرجاع المعلومات ككل، وعلى ذلك يمكن تحديد مفاهيم الاستدعاء والتحقيق كالتالي:

أولاً : الاستدعاء : هو عبارة عن مجموعة الوثائق المسترجعة المتصلة بموضوع البحث، التي تم استدعاؤها من كل ما يتصل بموضوع البحث داخل النظام.

وقد عبر الشكل عن ذلك بأنه عندما تكون كل الوثائق المتصلة بالموضوع داخل النظام ككل هي (relevant documents (Rd، والوثائق المتصلة بالموضوع المسترجعة منها هي (retrieved relevant documents (Rr ، فإن المعادلة المعبرة عن معامل الاستدعاء تتمثل في :

الاستدعاء = مجموعة الوثائق المتصلة بالموضوع والمسترجعة من مجموعة كل الوثائق المتصلة بالموضوع.

الاستدعاء = Rd \ Rr

ثانيًا : التحقيق : هو عبارة عن مجموعة الوثائق المسترجعة المتصلة بموضوع البحث من بين كل ما تم استرجاعه من قبل النظام.

وقد عبر الشكل عن ذلك بأنه عندما تكون كل الوثائق المسترجعة من النظام ككل هي (response set (Rs ، والوثائق المتصلة بالموضوع المسترجعة منها هي retrieved relevant documents (Rr) ، فإن المعادلة المعبرة عن معامل التحقيق تتمثل في :

التحقيق = مجموعة الوثائق المتصلة بالموضوع المسترجعة من مجموعة كل الوثائق المسترجعة.

التحقيق = Rs \ Rr

تدل التعريفات السابقة لكل من الاستدعاء والتحقيق على التطابق في أنهما معًا يعبران عن مجموعة الوثائق المتصلة بموضوع البحث المسترجعة على واجهة النتائج أمام المستفيد. غير أن كل منهما ينظر إلى هذه الوثائق المتصلة المسترجعة من منظور خاص به ؛ **فالتحقيق** يقيس نسبة الوثائق المتصلة بالموضوع المسترجعة تبعًا لكل ما تم استرجاعه، دون الاهتمام بالوثائق المتصلة بموضوع البحث التي كان يفترض استرجاعها ولم يسترجعها النظام. وهذه هي وجهة نظر المستفيد الذي يهتم فقط بما تقع عليه عيناه من معلومات مسترجعة، وما يهمه هو أن تكون كل المعلومات المسترجعة تفيده وتلبي حاجاته وعلى ذلك يقيس نسبة التحقيق.

أما **الاستدعاء** فيقيس نسبة الوثائق المتصلة بالموضوع المسترجعة تبعًا لكل ما يتصل بموضوع البحث داخل النظام، وعلى ذلك فإذا كانت هذه الوثائق

المسترجعة هي بالفعل كل ما يخص الموضوع، كان الاستدعاء في أفضل حالاته، أما إذا أخطأ النظام في استرجاع المتصل بالموضوع كان الاستدعاء في أدنى حالاته.

لقد شهد تاريخ نظم استرجاع المعلومات علاقة نمطية تربط بين الاستدعاء والتحقيق؛ حيث تؤكد دائما دراسات نظم الاسترجاع عن وجود علاقة عكسية بين مقياسي الاستدعاء والتحقيق، فكلما زاد الاستدعاء قل التحقيق، والعكس من ذلك. غير أن الباحث يرى أن هذه العلاقة تنطوي على رؤى أخرى يجب أن توضع في الاعتبار خاصة بعد ظهور نظام استرجاع المعلومات الكوني (الويب). ومن هذه الرؤى: -

أولاً: إن من يحكم على نسبتي الاستدعاء والتحقيق عند عرض النتائج هو المستفيد، وليس النظام. ومن ثم فإن هذا الحكم يتوقف على الشكل والصورة التي تكون عليها احتياجات المستفيد، التي قد تكون غير محددة أو مشوشة بما يؤثر على قرار وتقييم المستفيد لنتائج البحث.

ثانيًا : إن النتائج المسترجعة إنما هي الوجه الآخر من إستراتيجية البحث والكلمات المفتاحية، وهذه جميعًا تصاغ بإرادة المستفيد واختياره للكلمات المعبرة عن الموضوع. فما كان من استدعاء فهو منها، وما كان من تحقيق فهو بها.

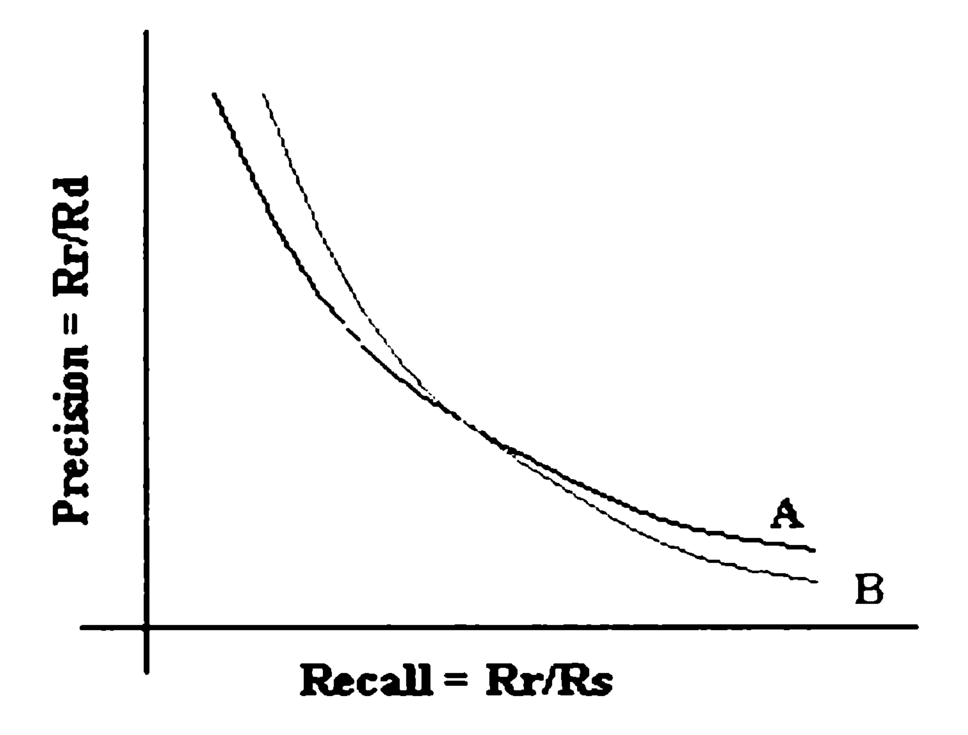
ثالثًا: أوجدت نظم استرجاع المعلومات على الويب منظورًا جديدًا لمقياسي الاستدعاء والتحقيق، وخاصة الاستدعاء؛ حيث يملك أي من نظم استرجاع المعلومات على الويب (مثل google) ما يقدر بآلاف الملايين من الوثائق في

شكل صفحات ومواقع، وعلى ذلك فإن ثمة بحثًا بسيطًا مع استخدام كلمات مختلفة يمكنه استدعاء الآلاف من الوثائق، وهو حجم لا يمكن لمستفيد فحصه أو التجول داخله بما يفقده الحكم الدقيق على الاستدعاء.

رابعًا : إن حساب معامل الاستدعاء كان أمرًا ممكنًا مع نظم استرجاع المعلومات التقليدية أو على الخط المباشر، أما في حالة نظم استرجاع معلومات الويب فهو أمر غير ممكن بحال من الأحوال. ويتضح هذا الأمر من أن الاستدعاء هو عدد الوثائق المسترجع المتصل بالموضوع من العدد الكلي لما يتصل بالموضوع من وثائق داخل النظام، وهو أمر كان يمكن حسابه مع النظم الثابتة أو بطيئة التحديث والإضافة، أما في حالة محركات البحث كنظم استرجاع على الويب فلا يمكن تحديد أو حتى تخيل عدد الوثائق المتصلة بموضوع ما داخل قاعدة بيانات النظام، ومن ثم لا أستطيع القول بأن ما تم استدعاؤه كثير أو قليل من كل الوثائق المتصلة بالموضوع داخل النظام.

خامسًا: إنه من الممكن تحويل العلاقة بين الاستدعاء والتحقيق من علاقة عكسية إلى علاقة طردية، وهذا يأتي من خلال اعتماد منهج دقيق ولغة محكمة في أثناء عمليات التكشيف، ثم القيام بانتقاء كلمات البحث والدقة في صياغة إستراتيجيات البحث، وبذلك يمكن استرجاع كل الوثائق التي تخص الموضوع، والابتعاد عن الوثائق غير المتصلة به، وهو ما يعني أعلى نسبة استدعاء وأعلى نسبة تحقيق.

يعبر الشكل رقم (٦ -٤) عن العلاقة النمطية التي تربط بين الاستدعاء والتحقيق، التي تلائم نظم استرجاع المعلومات الأخرى دون نظم استرجاع الويب.



الشكل رقم (٦ -٤) العلاقة العكسية بين الاستدعاء والتحقيق

إن استخدام مُعَامِلات الاستدعاء أو التحقيق لقياس كفاية نظام الاسترجاع، لا يعبر نجاحهما عن نجاح نظام استرجاع المعلومات في تلبية احتياجات المستفيدين من المعلومات وانتفاعهم بها ؛ حيث توجد عوامل خارجية لا صلة لها بنظام الاسترجاع تؤثر على تحقيق الاستفادة من المعلومات التي يحصل عليها ،ومن هذه العوامل ما يلي[1] :

---

(١) طلال ناظم الزهيري. **العوامل المتصلة بنظام استرجاع المعلومات وتأثيرها على كفاية الاسترجاع** .- اسم المدونة.- تاريخ الاستشهاد ٢٠/١١/٢٠٠٧، مسار http://azuhairi.jeeran.com/files/81729.doc

١) عدم خبرة الفرد السابقة بالموضوع الذي يقرؤه، وبالتالي عدم فهمه واستيعابه للمعلومات.

٢) عدم معرفة الفرد باللغة المكتوب بها المعلومات .. الأمر الذي يحول دون فهمه للمعلومات.

٣) عدم معرفة الفرد بالقراءة .. الأمر الذي يجعل حصوله على المعلومة غير كاف في حد ذاته لتحصيل المعرفة.

وخلاصة القول إنه لا يمكن الاعتماد في تقييم نظام استرجاع المعلومات فقط على معاملي الاستدعاء والتحقيق، لما لعملية بث المعلومات والتفاعل بين النظام والمستفيد من عوامل أخرى قد تعود إلى كفاية المستفيد ذاته وقدرته على الاستفادة من المعلومات المسترجعة من النظام.

٦ / ١ / ٤ / ٢ الفقد:

يعبر الفقد في تقييم نظم استرجاع المعلومات عن مجموعة الوثائق المتصلة بموضوع البحث غير المسترجعة أمام المستفيد. وبذلك يمكن القول إن الفقد هو الشق الثاني مع الاستدعاء في مجموعة الوثائق المتعلقة بموضوع البحث. فإذا كان عدد الوثائق التي تتصل بموضوع ما هو مئة، وكان المسترجع منها ستين وثيقة، فإننا نطلق على الأربعين وثيقة غير المسترجعة من النظام مصطلح الفاقد في نظام الاسترجاع. ومن ثم يمكن التعبير عن مُعَامِل الفقد بالمعادلة التالية:

معامل الفقد = مجموعة الوثائق المتصلة بالموضوع ككل (-) مطروحا منها مجموعة الوثائق المتصلة المسترجعة.

وتبعًا للشكل رقم (٦-٣) فإن معامل الفقد هو :

الفقد = Rd - Rr

ويبقى القول بأن مُعَامل الفقد هو معامل تقديري يمكن الوصول إلى نسبة أو عدد يمثله فعليًا في نظم استرجاع المعلومات التقليدية أو على الخط المباشر، أما في حالة نظم استرجاع الويب، فإن معامل الفقد هو معامل وهمي تخيلي لا يمكن الحصول على تقدير له أو عدد يمثله ؛ ويأتي ذلك من الديناميكية والتغير السريع في محتوى قواعد بيانات نظم استرجاع الويب، فإننا عند إجراء بحث عن موضوع ما داخل أحد محركات البحث لا يمكن بحال من الأحوال تخيل حجم المعلومات أو عدد الوثائق المتصلة بهذا الموضوع وقت تنفيذ البحث، فالإضافة والتحديث والتغيير عمليات تتم على مدار الدقائق التي يجرى فيها بحث نظام استرجاع الويب.

٦ / ١ / ٤ / ٣ الشوشرة :

يعبر مُعَامل الشوشرة عن مجموعة الوثائق المسترجعة غير المتصلة بموضوع البحث. وقد سميت شوشرة؛ لأنها تأتي ضمن النتائج المسترجعة بلا فائدة تحققها، فضلا عن كونها تعيق المستفيد عن تصفح النتائج المتصلة بالموضوع. ودائمًا ما تكون العلاقة بين التحقيق والشوشرة عكسية في حجم وأعداد الوثائق المسترجعة ؛ فإذا ما نجح النظام في الاقتصار على استرجاع أكبر عدد من الوثائق المتصلة بالموضوع فقط، فهو بذلك قد نجح في خفض نسبة الشوشرة المسترجعة. ويتم التعبير عن معامل الشوشرة بأنه:

مُعَامل الشوشرة = مجموعة الوثائق المسترجعة المتصلة بالموضوع (-) مطروحًا منها مجموعة الوثائق المسترجعة غير المتصلة بالموضوع.

وتبعًا للشكل رقم (٦-٣) فإن مُعَامل الشوشرة هو:

الشوشرة = Rr – Rs

ونهاية، فإن المعاملات الأربعة السابقة ترتبط جميعًا بالشكل الذي تجرى عليه عمليات التكشيف والتنظيم، فلا يمكن النظر لاسترجاع قدر معين أو استبعاده من الوثائق المتصلة بموضوع البحث دون الانتباه إلى تأثير عمليات التنظيم في استرجاع واستدعاء هذه الوثائق مرة أخرى.

٦ / ١ / ٤ / ٤ الترتيب و منزلة الوثائق :

تزداد أهمية الترتيب والتنظيم للوثائق المسترجعة من النظام كلما زاد حجم قاعدة البيانات، وزادت أعداد الوثائق المتوقع استرجاعها عن البحث بموضوع ما. وثمة جانب تعمل نظم استرجاع المعلومات على إبرازه عند استرجاع وثائق محددة، ألا وهو التفاوت في أهمية هذه الوثائق لتلبية الاحتياجات الموضوعية، ومن ثم توضع الوثائق ذات الأهمية الأكبر في مقدمة الترتيب ثم تليها الوثائق الأقل فالأقل أهمية. وتتخذ نظم استرجاع المعلومات إجراءين لترتيب قائمة الوثائق المسترجعة ؛ الأول : تحديد منزلة وأهمية كل وثيقة مسترجعة بالنسبة لموضوع البحث relevance. الثاني : ترتيب هذه الوثائق ترتيبًا تنازليًا اعتمادًا على منزلة وأهمية كل وثيقة في تقديم المعلومات. وتضع نظم استرجاع الويب أسسًا مختلفة لتحديد منزلة الوثائق المسترجعة وترتيبها في عرض النتائج، ومنها[1]:

---

(١) Results Ranking. Net Industries, 2007, cited 20/11/2007, cited at http://ecommerce.hostip.info/pages/912/Results-Ranking.html

١- تقدر أهمية الوثائق اعتمادًا على مرات التكرار التي وردت بها كلمات البحث في هذه الوثائق. ومن ثم فإن الوثائق الأكثر أهمية هي صاحبة أعلى تكرار لكلمات البحث.

٢- تقدر أهمية وقيمة الوثيقة بحسب المواقع التي وردت فيها كلمات البحث داخل جسم الوثيقة، ومن ثم فإن الوثائق ذات الأهمية الأكبر هي التي تحمل كلمات البحث في العنوان، ثم تليها الوثائق التي احتوت كلمات البحث في المقدمة، وتليها التي احتوت كلمات البحث في المتن... الخ.

٣- تقدر أهمية الوثائق تبعًا لعدد كلمات البحث الواردة بها، فإذا ما تكونت استراتيجية البحث من خمس كلمات كانت الوثائق الأكثر أهمية هي التي تحتوي على خمس كلمات، ثم تليها الوثائق التي تحتوي على أربع كلمات من مجموع كلمات البحث، وتستمر هذه السلسلة وصولاً إلى الوثائق التي تحتوي على كلمة واحدة من كلمات البحث.

٤- تقدر أهمية الوثائق اعتمادًا على مدى القرب والبعد بين كلمات البحث في متن الوثيقة، فإذا ما جاءت كلمات البحث متتابعة بلا فواصل من كلمات أخرى كانت هذه الوثيقة هي الأكثر أهمية، ثم تليها الوثائق التي تحوي كلمات البحث مبتعدة عن بعضها في المتن، وكلما زاد الفاصل والبعد بين كلمات البحث، قلت أهمية الوثيقة في الصلة بموضوع البحث.

٥- تقدر أهمية الوثائق بحسب الروابط أو الصلات بين الوثيقة والوثائق الأخرى المرتبطة بموضوع البحث.

٦- عدد الروابط غير النشطة داخل صفحة الويب مما ينقص من قيمة الترتيب.

٧- تقييم لغة الترميز المعيارية داخل الصفحة، وحصر أخطاء الرموز والتصميم.

٨- مقدار وقت التحميل الذي تقل القيمة معه كلما زاد وقت تحميل الصفحة.

ويعد عامل الترتيب بين النتائج المسترجعة أحد أهم الملامح المؤثرة في تقييم المستفيد لنظم استرجاع المعلومات ؛ حيث يصعب على المستفيد استعراض النتائج المسترجعة غير المرتبة، خاصة إذا ما بلغت أعداد هذه النتائج عشرات الآلاف من الوحدات. ومن ثم كان نجاح نظام الاسترجاع في ترتيب الوثائق أحد المُعَاملات الإيجابية في الحكم على كفاية النظام.

٦ / ١ / ٤ / ٥ خصائص العرض للوسائط المتعددة :

تحتاج بعض نظم استرجاع المعلومات إلى مقاييس أو معاملات إضافية خلاف الوارد ذكرها، وهذا يعود إلى اختلاف طبيعة النظم أو طبيعة الوثائق التي تحتويها. وينطبق هذا الأمر على نظم استرجاع الوسائط المتعددة مثل الصور الثابتة والمتحركة أو المواد الرسومية الأخرى ؛ وتحتاج هذه النظم إلى مزيد من الاهتمام بآليات بحث الوسائط، بجانب خصائص عرضها كنتائج مسترجعة، فلابد من تميز واجهة عرض نتائج الوسائط بما يجعل المستفيد

قادرًا على التصفح اليسير للنتائج، فتعرض اللقطات الصغيرة إذا ما كانت النتائج صورًا فوتوغرافية، أو مشاهد متحركة إذا ما كانت النتائج مواد فلمية. وهناك الكثير من الجوانب التي يجب الاهتمام بها في واجهات عرض الوسائط المتعددة، وقد تطرق إليها الباحث في دراسة سابقة عن محركات بحث الصور الفوتوغرافية، ومن هذه الجوانب:

شكل الوسائط المعروضة :

تحتاج نظم استرجاع الوسائط المتعددة إلى صياغة ملامح أخرى غير التي تتعلق بتنظيم وبحث الوسائط المتعددة، وتتعلق هذه الملامح بالجانب المرئي الوسائط المتعددة ، ومن هذه الملامح؛ اختيار الوسائط المتعددة الرقمية ذات الكثافة الملائمة لعرض محتوى الوسائط المتعددة Resolution، وأيضًا حجم اللقطات المعروضة داخل النتائج، بجانب التداخل بين البيانات واللقطات داخل وحدات النتائج المعروضة. وتأتي أهمية الشكل المعروض به النتائج لما له من تأثير على سرعة تصفح النتائج مع تحديد المستفيدين لمجموعات الدقة والشوشرة داخل النتائج.

البيانات الببليوجرافية المصاحبة :

تعتمد بعض نظم استرجاع الوسائط المتعددة على إلحاق بعض البيانات الببليوجرافية مع الوسائط المتعددة المعروضة في واجهات النتائج. ويمكن عرض اللقطات منفردة مع وضع رسومات لتحديد اختيارات التعامل، مع حذف البيانات الببليوجرافية على أن تكون التسجيلة الببليوجرافية متاحة أمام المستفيدين دون إجراء تعديلات في البيانات المعروضة.

عدد وحدات الوسائط المعروضة :

يرتبط عدد اللقطات أو الوحدات المعروضة في نتائج البحث مع سرعة تصفح المستفيد لهذه النتائج، على أن التحديد المسبق من جانب محرك البحث قد لا يلائم ما يرجوه المستفيد من سرعة المرور على مئات الوسائط المتعددة. و يمكن التغلب على عدد اللقطات بفتح الكثير من النوافذ أمام المستفيد إذا ما حجب محرك البحث عدد اللقطات المعروضة، كما يفعل الكثير من مستخدمي الويب للتغلب على بطء التحميل.

الروابط داخل عرض النتائج :

تعمل محركات بحث الوسائط على إلحاق مجموعة من الروابط مع اللقطات الممثلة للوحدات المسترجعة، وتشمل هذه الروابط الانتقال إلى الصفحة المصدرية للارتباط بصفحة الحجم الحقيقي للصورة مثالاً والذي ينتهي مساره دائما بـ GIF, JP. ويمكن استبدال هذه الروابط بمجموعة من الرسومات النشطة المعبرة عن اختيارات التعامل، تشير إحداها إلى التسجيلة الكاملة للصورة، أو الانتقال للصفحة المصدرية، أو تحميل الوسائط المتعددة، أو إرسالها كرسالة بريد إلكتروني، أو عرض الوسائط المتعددة فقط في الحجم الحقيقي.

٦ / ٢ نماذج قياس كفاية نظم استرجاع المعلومات :

لقد شكل هذا الفصل مقترحًا لوضع نموذج عام يحتوي على مختلف معاملات القياس والتقييم التي يمكن من خلالها النظر إلى نظم استرجاع المعلومات، سواء أكانت التقليدية وعلى الخط المباشر أم تلك التي تعمل في

فضاء الويب. لقد سجلت الدراسات السابقة لنظم الاسترجاع نماذج لقياس كفاية النظم، غير أنها تقتصر على معاملات الاستدعاء والتحقيق في تقييم أداء نظم الاسترجاع.

ويشتمل الجزء التالي دراسة بعض النماذج المطروحة؛ لتقييم نظم الاسترجاع، وقد قسمها الباحث إلى ما يلائم نظم استرجاع المعلومات الثابتة أو التقليدية متمثلة في نظامي كنت ولانكستر، ونظم الاسترجاع الديناميكية أو على الويب، ويمثلها نموذج قد طرحه الباحث.

٦ / ٢ / ١ نماذج لقياس كفاية نظم الاسترجاع الثابتة :

٦ / ٢ / ١ / ١ نموذج ألن كنت :

تعتمد فكرة هذا القانون على أن أي نظام لاسترجاع المعلومات يحتوي على عدد من الوثائق تمثل في مجموعها " ن " وأن هناك عدد " م " وثيقة يحتمل أنها تقع في مجال اهتمام المستفيد. وأن هناك عددًا أقل من الوثائق " ف " يمثل الاحتياج الفعلي للمستفيد. وعلى ذلك تكون[1]:

ن = عدد جميع الوثائق التي يضمها النظام.

م = عدد الوثائق التي يحتمل أنها تقع في مجال اهتمام المستفيد .

ف = عدد الوثائق التي تمثل الاحتياجات الفعلية للمستفيد .

ومن ثم تكون العلاقات كما يلي :

(١) هاني محيي الدين عطية. **محاضرات غير منشورة.**

معامل الكثافة = نسبة الوثائق التي تقع في مجال اهتمام المستفيد = عدد الوثائق المحتمل \ عدد الوثائق الكلي = **م \ ن**.

معامل الاستبعاد = نسبة الوثائق المستبعدة = **ن - م \ ن**.

معامل الاستبقاء = نسبة الوثائق التي تمثل احتياجات المستفيد الفعلية = **ف \ م**.

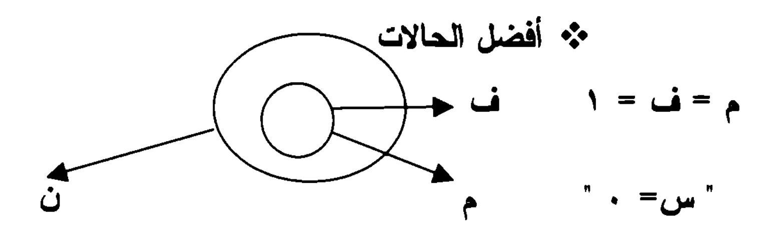
معامل التشويش= نسبة الوثائق التي تم استرجاعها ولا تقع ضمن احتياج المستفيد = **م – ف \م**.

وحيث إنه لا يوجد نظام استرجاع نموذجي، فإنه يظل هناك احتمال دائم أن هناك عددًا من الوثائق تقع في مجال اهتمام المستفيد و لم يتم استرجاعها و هي " س " وتسمى معامل الفقد.

معامل الفقد = نسبة الوثائق التي تقع في مجال اهتمام المستفيد ولم يتم استرجاعها = **س \ ن** .

ويمكن تمثيل الحالات المختلفة كما يلي :

❖ **أفضل الحالات**

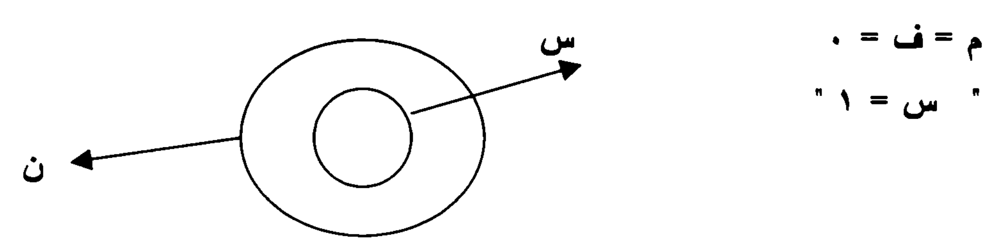


الشكل رقم (٦-٥ أ) أفضل الحالات لنظام الاسترجاع

❖ اسوأ الحالات

م = ف = ٠

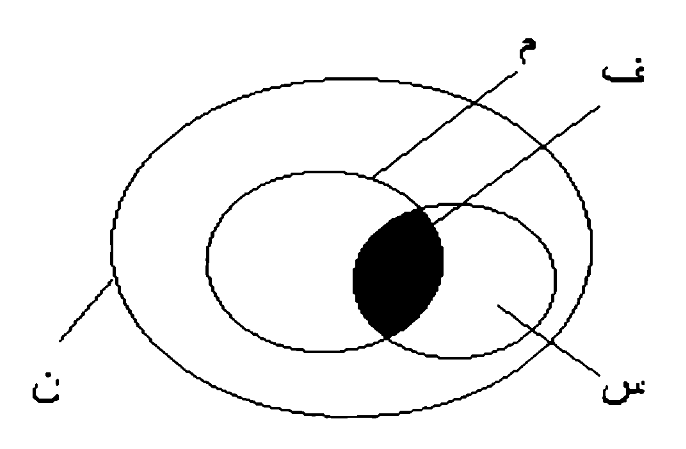
" س = ١ "



الشكل رقم (٦ -٥ ب) أفضل الحالات لنظام الاسترجاع

❖ الحالة العادية

٠ < س < ١ ٠ < ف < م



الشكل رقم (٦ -٥ ج) أفضل الحالات لنظام الاسترجاع

٦ / ٢ / ١ / ٢ نموذج لانكستر:

لقد بدأ لانكستر بالحديث عن مقياسي الزمن والتكلفة وهما يعتبران وجهي عملة استخدام الويب ؛ حيث يترجم الوقت المبذول في البحث إلى تكلفة من جانب الباحث أو تكلفة في الوقت، ولا تختلف هذه المقاييس بين جميع

نظم وخدمات استرجاع المعلومات؛ إلا أن تقرير الكفاية يتوقف على الوحدات المسترجعة من جانب النظم لإفادة الباحثين؛ فإذا ما كانت هذه الوحدات تلبي الجانب الأكبر من احتياجات المستفيد أصبحت التكلفة في مستوى منخفض إلى جانب الوقت المبذول لاسترجاع تلك الوحدات، وتسمى الوحدات المسترجعة المطابقة لاحتياجات المستفيد بالتحقيق؛ أحد أهم عوامل الحكم على كفاية الاسترجاع. وبالتعرض إلى تفسير نظام استرجاع لانكستر للمعلومات يتضح أن مفهوم الاستدعاء الذي يعتمد عليه المستفيد في تحديد كفاية النظام يختلف عما يوضحه الجدول رقم (٦-١) لنظام معاملات الكفاية للانكستر؛ حيث ينظر المستفيد عند بحث نظم الاسترجاع التقليدية وعلى الويب، إلى أن الاستدعاء هو كل ما تم استرجاعه من داخل قاعدة البيانات ككل : أي كل المسترجع. على حين أن معامل الاستدعاء عند لانكستر يتم حسابه اعتمادًا على معرفة عدد الوحدات المسترجعة ذات الصلة بالموضوع (وهو ما لا يمكن إدراكه داخل محركات البحث سريعة التحديث)، وعلى ذلك فإن الاستدعاء عند لانكستر هو فقط مجموعة من كل المسترجع إلا أنه يساوي كل المسترجع عمليًا عند المستفيد. ويوضح الجدول رقم (٦-١) معاملات الكفاية عند لانكستر[١]:

(١) لانكستر، ولفرد. **نظم استرجاع المعلومات**؛ ترجمة حشمت قاسم.- القاهرة : دار غريب ، ١٩٨١م.

الجدول رقم (٦ -١) معاملات قياس الكفاية نموذج لانكستر الملائم لنظم الاسترجاع الثابتة

	صالح للموضوع أ + ج	غير صالح للموضوع ب + د
مسترجع أ + ب	أ استدعاء تحقيق	ب شوشرة
غير مسترجع ج + د	ج فاقد	د استبعاد

اعتماداً على الجدول رقم (٦-١) تكون معاملات قياس الكفاية كما يلي :

معامل الاستدعاء = الوحدات المسترجعة الصالحة \ كل الصالح للموضوع= **أ \ أ+ج**

معامل التحقيق =الوحدات المسترجعة الصالحة \ كل المسترجع = **أ \ أ + ب**

معامل الشوشرة = الوحدات المسترجعة غير الصالحة \ كل المسترجع = **ب \ أ + ب**

معمل الفاقد = الوحدات غير المسترجعة الصالحة \ كل الصالح = **ج \ أ+ج**

معامل الاستبعاد= الوحدات غير المسترجعة غير الصالحة \ كل غير المتصل= **د \ ج+د**

٦ / ٢ / ٢ نماذج قياس كفاية نظم الاسترجاع على الويب:

٦ / ٢ / ٢ / ١ نموذج سيد ربيع:

إن اعتماد نظم استرجاع المعلومات التقليدية على نظام لانكستر يختلف بدوره عما يمكن التعامل معه في نظم استرجاع الويب، وذلك إنما يرجع إلى

الديناميكية التي تعمل بها نظم استرجاع الويب في إضافة آلاف المواقع يوميا إلى قواعدها، فضلا عن اعتماد أساليب مختلفة في التنظيم والبحث لصفحات الويب وملفات الصور، إلى جانب البعد التام عن التدخل البشري في تنظيم وبحث محركات بحث الويب. ويجب ملاحظة وجود الاستدعاء في النوع السابق كجانب مكمل للفاقد لتكملة كل المتصل بالموضوع الذي لم يسترجع، في حين أن الناحية العملية تحتم إطلاق مصطلح الاستدعاء على كل الوحدات المسترجعة لكي تلائم التطبيق العملي لحكم المستفيدين. ويطرح الباحث النموذج التالي الممثل في الجدول رقم (٦-٢) للاستخدام في تحديد كفاية نظم استرجاع الويب عامة، كما يلي[١] :

الجدول رقم (٦ -٢) معاملات قياس الكفاية المقترحة لنظم استرجاع الويب

	صالح للموضوع أ + ج	غير صالح للموضوع ب + د
مسترجع (الاستدعاء) أ + ب	أ تحقيق	ب شوشرة
غير مسترجع د	د استبعاد	

وتكون المعاملات كما يلي :

معامل الاستدعاء = كل المسترجع \ كل ما يشمله النظام = أ + ب \ أ + ب + د

معامل التحقيق = كل المسترجع الصالح\ كل الاستدعاء (المسترجع) أ \ أ + ب

(١) سيد ربيع سيد إبراهيم. **محركات بحث الصور الثابتة على الإنترنت** : دراسة تحليلية .- الرياض : مكتبة الملك فهد الوطنية، ٢٠٠٧.

معامل الشوشرة= كل المسترجع غير الصالح\ كل الاستدعاء (المسترجع) ب \أ + ب

معامل الاستبعاد =كل المستبعد \ كل ما يشمله النظام = د \أ + ب + د

تم بناء النظام السابق على أسس هي :

- إن من يصنع التنظيم داخل محركات البحث هو برنامج (نظام المعلومات) ومن غير المؤكد معرفة عدد الوحدات التي تلبي موضوع المستفيد.
- إن استخدام مصطلح الفقد مع الاسترجاع في النظم الآلية لا يوافق بيئة الويب التي تعمل على تكشيف كل كلمات النص التي تصبح نقاط استرجاع للنص، ومن ثم فإن التحكم في استدعاء النصوص يرجع إلى المستفيد المحدد للموضوع. واعتمادًا على عدم تحديد النظام لما يحويه من وحدات عن موضوع بعينه، فإن حساب الفقد أو الفاقد من الوحدات ليس له معنى.
- إن طبيعة النظم الآلية تبنى على استدعاء كل ما يطابق إستراتيجية البحث ومن ثم لا يعطي النظام من جانبه أي شوشرة، وإنما يحدد المستفيد تلك الشوشرة من وجهة نظر احتياجات خاصة له. حيث يسترجع النظام كل ما تستدعيه كلمات البحث من وحدات.
- إن حقيقة التعامل مع نظم المعلومات الآلية هي أن الاستدعاء هو كل المسترجع وأن التحقيق هو ما يطابق احتياج المستفيد من داخل الاستدعاء.

ويتعامل النموذج السابق مع طبيعة نظم استرجاع المعلومات على الويب التي تتعامل مع الموضوعات في صفحات الويب كما قام عليه مصممو هذه الصفحات. حيث ينتج عن تكشيف محركات البحث للنص الكامل أن تكون مختلف كلمات النص مداخل وكلمات مفتاحية يتم الاسترجاع من خلالها.

٦ / ٣ الخاتمة:

لقد تغيرت ملامح نظم استرجاع المعلومات بالشكل الذي يجعلنا نحتكم إلى آليات تقييم وقياس الكفاية خلاف ما كان يستخدم من معاملات ومعايير في قياس الكفاية. ولقد قدم الباحث في هذا الفصل طرحًا لإطار شامل لتقييم وقياس كفاية نظم استرجاع المعلومات ؛ وذلك من خلال الوقوف على الجوانب العامة التي يقاس من خلالها أداء النظام ثم العناصر الفرعية المكونة للإطار العام. وكانت الجوانب العامة للتقييم وقياس الكفاية تتمثل في جوانب تخص النظام ذاته، وجوانب تخص المستفيد، وجوانب تخص النتائج المسترجعة، وجوانب تخص العمليات الداخلية للنظام. ثم تم تناول كل واحد من هذه الجوانب مع تتابع عناصره بالتفصيل.

خلاصة البحث:

خرج الباحث من هذه الدراسة بمجموعة من النتائج هي:

١- إن علم المعلومات بصفة عامة ونظم استرجاع المعلومات بصفة خاصة، قد شهدت تغييرات سريعة متلاحقة طرأت عليهما منذ بداية استخدام الحاسب الآلي، حتى قدمت الويب لتتوج هذه التغييرات باستحداث ملامح وخصائص جديدة لمفاهيم وأداء وواقع نظم استرجاع المعلومات الرقمية خاصة دون التقليدية. حيث أثرت الويب في إعادة النظر إلى مفهوم نظام استرجاع المعلومات، والخصائص التفاعلية لهذه النظم، فضلاً عن مجموعة الأشكال المستحدثة من المعلومات الصوتية والمصورة التي ملأت الويب.

٢- إن تأثير علم المعلومات على علم المكتبات قد اتضح جليًا من أدبيات الإنتاج الفكري الحديث ؛ حيث اتجهت مختلف دراسات المكتبات إلى التحولات الرقمية ودراسة آثار استخدام الرقمنة في أنشطة المكتبات، وهو ما يعني لدى الباحث التغييرات الشكلية في علم المكتبات مفهومًا وأداء حتى أن وصلت مع الويب كنظام رقمي كوني لاسترجاع المعلومات إلى جوهر أنشطة المكتبات؛ لنشهد تغييرًا في المفاهيم مثل استدعاء أم استرجاع، تنظم أم معالجة فنية، لغات تنظيم أم أساليب استرجاع... وغيرها من المفاهيم.

٣- إن حقيقة لا يمكن إنكارها هي أهمية التفاعل الإيجابي من جانب المستفيدين ؛ حيث تهتم دراسات التقييم وقياس الأداء في نظم استرجاع المعلومات بالتغذية المرتدة أي رأي المستفيد، وهو الرأي المتمثل في عوامل الاستدعاء والتحقيق. ومن ثم يرى الباحث أنهما معياران يخضعان فقط

لوجهة نظر المستفيد غير المحددة أو التي يمكن قياسها بأساليب كمية وإحصائية توضح بالفعل مستوى كفاية نظام الاسترجاع.

٤- إن ثمة جانبًا مهمًا في معاملي قياس الكفاية : الاستدعاء والتحقيق، لم تلفت إليه أيٌّ من دراسات التقييم لنظم استرجاع المعلومات ؛ هو أن كلا من الاستدعاء والتحقيق ليسا بالكفاية أو الفاعلية اللازمة للحكم على كفاية وأداء نظام الاسترجاع، ويأتي ذلك من كونهما انعكاسًا للعمليات الفنية التي تتم على الوثائق داخل النظام، حيث يزداد الاستدعاء إذا ما اعتمد النظام على مبدأ الإحاطة والشمول في عمليات التكشيف، ويزداد التحقيق إذا ما انحاز النظام لمبدأ التخصيص والتعمق في تحليل محتوى الوثائق. فكثرة الكلمات الدالة تعني استدعاء أكبر وتحقيقًا أقل، وقلتها تعني العكس. ومن ثم فإنه لا يجب الأخذ بمعاملي الاستدعاء والتحقيق في الحكم على كفاية النظام قبل النظر إلى طبيعة أداء العمليات الفنية داخل النظام.

٥- إن مفهوم نظام استرجاع المعلومات لم يرتبط بظهور المكتبات التقليدية أو نظم الاختزان الرقمية، بل يمكن القول إن مفهوم نظام استرجاع المعلومات يمتد إلى أي نشاط من شأنه اختزان المعلومات مع مراعاة تنظيمها والرجوع إليها مرة أخرى، فالذاكرة البشرية هي نظام لاسترجاع المعلومات، والقوائم الببليوجرافية الورقية أو الرقمية هي نظم لاسترجاع المعلومات، ورموز ونقوش الحضارات القديمة هي نظم لاسترجاع المعلومات. وقد استمد الباحث هذه النتيجة من الأثر الذي أحدثه الحاسب الآلي في معالجة وتنظيم المعلومات، ثم ما أكدته الويب بخصائصها التنظيمية والاسترجاعية المختلفة للمعلومات الرقمية.

٦- إن التفاعل بين نظام استرجاع المعلومات والمستفيد لا يتوقف فقط على تغطية المعلومات أو طرق التنظيم والمعالجة داخل النظام، وإنما يتأثر أيضًا بخصائص الأداء والتفاعل الداخلي بين مكونات وعناصر النظام ؛ حيث تتميز الويب بخصائص التفاعلية بين أدواتها وبين المستفيد من ناحية، وديناميكية التغيير والتحديث والتطوير المتلاحق في شكل ومحتوى المعلومات من ناحية أخرى. وهو ما يمكن أن يدفع نظام استرجاع المعلومات إلى توفير أكبر قدر من الكفاية في الوصول لتلبية احتياجات المستفيدين.

٧- إن اختلاف نظم استرجاع المعلومات في البيئة التقليدية عنها في بيئة الخط المباشر عنها في بيئة الويب، قد أوجب الاحتكام إلى معايير ومعاملات للتقييم ولقياس الكفاية تختلف بحسب بيئة النظام. فقد أظهر الفصل السادس مدى الاختلاف بين نموذجي قياس الكفاية في النظم الثابتة أو التقليدية (نموذج لانكستر) عنه في النظم التفاعلية (نموذج سيد ربيع). ومن ثم فإن التطوير الدائم في الأداء لنظم الاسترجاع يستوجب معه التغيير المستمر في معايير ومقاييس تقييم النظم.

٨- إنه يمكن القول إن الويب قد حملت إلى نظم الاسترجاع أشكالاً رقمية خالصة من نظم الاسترجاع لا تعتمد في بنيتها أو معالجتها أو أساليب استرجاعها على معايير علم المعلومات، وإنما تعتمد على معايير وتقنيات علم الحاسب الآلي بصفة خاصة. وقد اتضح ذلك من ظهور ما يسمى بنظم الاسترجاع المبنية على المحتوى Content Based Retrieval Systems CBRS، وهذه النظم مثل استرجاع الصور ومواد المعلومات الرقمية المصورة اعتمادًا على خصائص تحليل اللون Color والبنية Texture الشكل Shape التي

تمثل خصائص رقمية لملفات المعلومات. ومن ثم فإن النظام هنا لا يعتمد على الكلمات الدالة أو التصفح أو الربط بالمنطق البولياني أو غيره، وإنما يعتمد على تقنيات التحليل الرقمي.

٩- إن الويب كنظام استرجاع للمعلومات قد قدمت خصائص للمستفيدين لم تقدمها نظم الاسترجاع في المراحل الزمنية السابقة لها مثل :

أولاً : إمكانية وصول المستفيدين إلى نظم الاسترجاع عن بعد من خلال الشبكة العالمية وإجراء عمليات البحث والاسترجاع للمعلومات التي تلبي الحاجات الموضوعية.

ثانيًا : إمكانية الربط بين التسجيلات الببليوجرافية التي تعطي مجرد إشارات عن المعلومات الأولية، بالنصوص الكاملة لها ومن ثم يستطيع المستفيد الوصول إلى المعلومات الأولية التي يحتاجها دون التوقف عند الإشارات الببليوجرافية.

ثالثًا : عملت الويب على استحداث مجموعة من نظم استرجاع المعلومات في أشكال ووسائط حمل معلومات لم تكن البيئة التقليدية أو حتى الإلكترونية تستطيع التعامل معها، مثل محركات بحث الصوت والصوت كنظم لاسترجاع الملفات الرقمية لوسائط الصوت والصورة، هذا إلى جانب الحديث عن نظم استرجاع قواعد البيانات ذاتها وإمكانية الجمع بين قواعد بيانات ببنية مختلفة واسترجاع المعلومات منها في وقت واحد.

١٠- إن ثمة حدًا فاصلاً يمكن إدراكه بين ما تقوم به نظم استرجاع المعلومات في استخدام أساليب علم المعلومات لأجل تنظيم واسترجاع

المعلومات، وبين ما تقوم به هذه النظم من استخدام أساليب تقنية حاسوبية خالصة لتنظم نفس المحتوى الرقمي. ومثالاً على ذلك نظم استرجاع المعلومات الرقمية المصورة الثابتة أو المتحركة ؛ حيث تستخدم هذه النظم أسلوبين لاسترجاع المعلومات هما استرجاع المعلومات المبني على النص TBCR، وذلك باستخدام البيانات عن البيانات. أو استخدام أسلوب الاسترجاع المبنى على المحتوى CBCR.

١١- إن نظم استرجاع المعلومات هي نظام من نظام أكبر ؛ حيث تتأثر نظم استرجاع المعلومات بمفاهيمها وإجراءاتها بمختلف التغييرات الرقمية والتقنية التي تحدث في البيئة الخارجة. وهو ما تم مناقشته من كون نظم استرجاع المعلومات دائمًا، سواء أكانت النظم التقليدية أم الرقمية هي نظم مفتوحة على البيئة الخارجية. ومن ثم كانت الويب وتقنياتها عوامل تأثير وتغيير في نظم استرجاع المعلومات علمًا وعملاً، حيث يمكن إضافة أو تغيير خصائص من نظم استرجاع المعلومات وفقًا لأي تغييرات تطرأ على عناصر بناء هذه النظم، مثل الوثيقة أو المستفيد أو اللغة التي يعمل بها نظام استرجاع المعلومات.

١٢- تتميز نظم الاسترجاع في البيئة الرقمية عن نظيرتها في البيئة التقليدية بإمكانية معالجة مختلف أشكال المعلومات ووسائط حملها، مما جعل المستفيدين على حرص أكبر في الاتجاه إلى نظم الويب، وهو خلاف ما تتقيد به النظم التقليدية في التعامل مع أشكال محددة من مصادر المعلومات ينحصر أغلبها في المواد النصية الورقية. ومن ثم يمكن القول إن نظم الويب يسرت أكثر سبل الوصول إلى المعلومات بأسرع وأدق وبأعلى درجات الاستدعاء عنها في النظم الأخرى.

١٣- إن نظم استرجاع الويب تعتمد على أتمتة عملياتها Automation مستخدمة مجموعة من البرامج الفرعية مثل الزاحف لإضافة المعلومات، والمكشف لتحليل المحتوى، وآليات البحث لاستدعاء المعلومات بحسب المصطلحات الموضوعية. وهذه الأتمتة هي ما تفتقده نظم الاسترجاع الأخرى. وحري بالذكر أن أتمتة العمل في نظم الويب تمثل في التحديث المستمر المتلاحق لمصادر الويب، تعدد هذه النظم بما جعلها تتجه إلى التخصص الدقيق، وأخيرًا الإتاحة والوصول إلى كم هائل من المعلومات، مما يعني وفرة الاستدعاء، مع إعطاء المستفيد أداة التحكم في التحقيق باختياره وانتقائه لكلمات البحث عن موضوعه.

١٤- تعمل الآن نظم استرجاع المعلومات في بيئة تعتمد أكثر ما تعتمد على صياغة الوثائق في المجال الموضوعي بألفاظ ومصطلحات مختلفة قد يستخدمها أو لا يستخدمها المستفيد عند استرجاع المعلومات.

١٥- إن نظم استرجاع المعلومات على الويب تستخدم لغة التكشيف الطبيعية المعتمد على التكشيف بالاشتقاق كأسلوب لتنظيم الوثائق، ومن ثم تصبح مختلف كلمات الوثيقة القريبة للموضوع والبعيدة عنه مداخل تستدعي هذه الوثيقة في حالة البحث بأي من هذه الكلمات.

١٦- إن معاملات الاستدعاء والتحقيق تتسم بالكفاية ودقة التطبيق فقط في حالة الاستخدام مع نظم الاسترجاع التي تستخدم اللغة المقيدة المحكومة بقوائم رؤوس الموضوعات والمكانز. وهو ما يقل وينخفض تدريجيًا كلما تم الاتجاه إلى استخدام اللغات الطبيعية في التكشيف والاسترجاع.

١٧- إن بيئة الويب بيئة ديناميكية المحتوى وعندما يقدم المستفيد على بحث أي من نظم استرجاع المعلومات على الإنترنت فهو يتعامل مع صندوق أسود من المعلومات لا يستطيع تحديد ملامح الوثائق داخله.

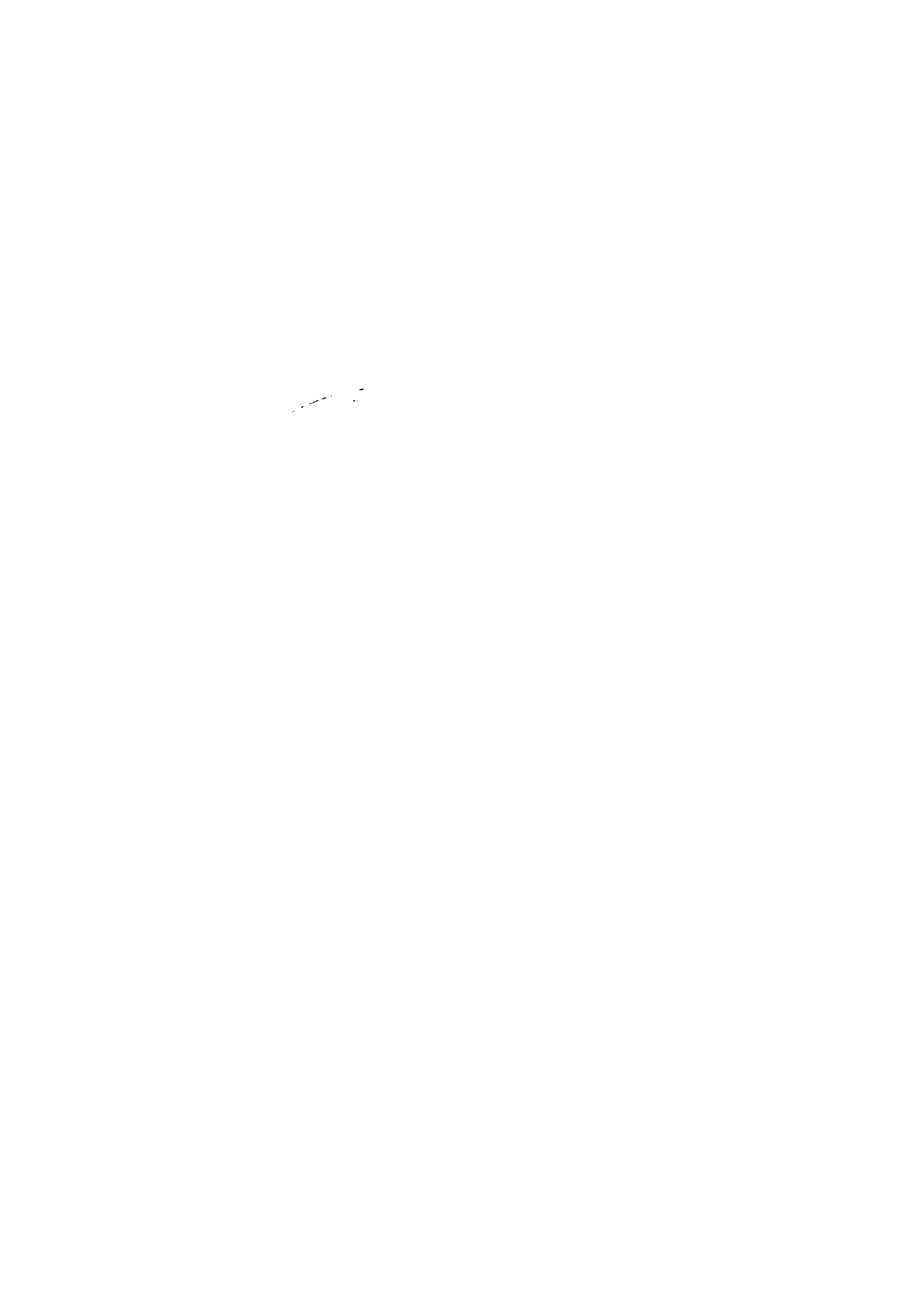
التوصيـــات :

يوصي الباحث بمجموعة من التوصيات هي:

١- ينبغي على أقسام المكتبات والمعلومات الالتفات إلى التأثير الرقمي وما يحدثه من تغيير في المفاهيم والأطر النظرية التي يسير بها علم المكتبات ؛ وذلك للعلاقة التي تربط بين علم المكتبات وعلم الحاسب الآلي من جذب وشد في استخدام التقنيات الرقمية ، فضلاً عن تكاملهما لبناء علم المعلومات.

٢- ينبغي على أساتذة المكتبات والمعلومات الانتباه نحو ما تسعى إليه تقنيات الحاسب الآلي والشبكات في بناء نظم استرجاع المعلومات التي تعتمد في معالجتها وتنظيمها على الحقول الببليوجرافية أو الميتاداتا metadata ، وإنما تعتمد كلية على تقنيات التحليل الرقمي مثل نظم الاسترجاع المبنية على المحتوى CBRS.

٣- ينبغي على طلاب وباحثي علم المكتبات والمعلومات الوقوف على حدود التغيير والتحديث الناتج عن ظهور الويب في فلك علم المعلومات، وثمة حد دقيق بين ما يمكن أن نتناوله من دراسات عن الويب من ناحية علم الحاسب والتقنيات الرقمية ، وبين ما يمكن أن نتناوله عن الويب من دراسات من ناحية علم المعلومات ، خاصة تنظيم وإدارة واسترجاع المعلومات الرقمية.



قائمة المصادر

أولاً : المصادر العربية :

- **المصادر العربية التقليدية :**

١- أحمد بدر و محمد فتحي عبد الهادي و ناريمان إسماعيل متولي . **التكشيف والاستخلاص : دراسات في التحليل الموضوعي** .- القاهرة: دار قباء، ٢٠٠١ .- ص٤٧ -٤٩.

٢- خالد محمد رياض. **أدلة ومحركات بحث شبكة الإنترنت : دراسة مقارنة**.- الاستراتيجية العربية الموحدة للمعلومات في عصر الإنترنت ودراسات أخرى .- تونس : الاتحاد العربي للمكتبات والمعلومات، (أكتوبر١٩٩٩). ص ١٣٩.

٣- سيد ربيع سيد إبراهيم. **محركات بحث الصور الثابتة على الإنترنت : دراسة تحليلية** .- الرياض : مكتبة الملك فهد الوطنية، ٢٠٠٧م.

٤- علي السليمان الصوينع. **استرجاع المعلومات في اللغة العربية** .- الرياض : مكتبة الملك فهد الوطنية، ١٩٩٤. ص- ص ٦٨-٧٧

٥- لانكستر، ولفرد. **نظم استرجاع المعلومات** ؛ ترجمة حشمت قاسم.- القاهرة : دار غريب ، ١٩٨١.

٦- هانئ محيى الدين عطية. **جهود تصنيف المعرفة في البيئة الإلكترونية** : دراسة استكشافية.....

٧- هانئ محيى الدين عطية. **محاضرات غير منشورة.**

- **المصادر العربية الرقمية :**

٨- أمل حسين عبد القادر. **المتطلبات الوظيفية للتسجيلة الببليوجرافية** .- cybrarians journal .- ع ٥ (يونيو ٢٠٠٥) .- تاريخ الاطلاع < اكتب هنا تاريخ اطلاعك على المقالة > .- متاح في:

http://www.cybrarians.info/journal/no5/frbr.htm

٩- **تعريف التكشيف ونظمه**.تاريخ الاستشهاد ٢٠٠٧/١١/٢٠، متاح على www.jeeran.com/archive/2007/3/168846.html

١٠- رجب عبد الحميد . **تقنيات الويب الدلالي للمكتبات الرقمية** .- cybrarians journal .- ع ١٤) سبتمبر ٢٠٠٧) . – تاريخ الإتاحة < اكتب هنا تاريخ اطلاعك على الصفحة . > – متاح في : http://www.cybrarians.info/journal/no14/semantic.htm

١١- طلال ناظم الزهيري. **العوامل المتصلة بنظام استرجاع المعلومات وتأثيرها على كفاءة الاسترجاع** ، تاريخ الاستشهاد ٢٠٠٧/١١/١١، مسار المصدر http://azuhairi.jeeran.com/files/81729.doc

١٢- طلال ناظم الزهيري. **دور المستفيد في كفاءة استرجاع المعلومات**، تاريخ الاستشهاد ٢٠٠٧/١١/١١، مسار المصدر

http://azuhairi.jeeran.com/files/81727.doc

١٣- علي بن شويش الشويش. **مارك MARC والبيانات الخلفية (الميتاداتا) Metadata: علاقة نديه أم تكاملية؟**، تاريه الاستشهاد ٢٠٠٧/١١/٢٠، مسار http://www.yamsyaf.com/informatics1.htm

١٤- فاتن سعيد بامفلح. **أساسيات نظم استرجاع المعلومات الإلكترونية** .- الرياض : مكتبة الملك فهد الوطنية، ٢٠٠٧م

http://www.kfnl.gov.sa/idarat/alnsherel/asaseiatnzmstrgaaalmalomat/pdf/f1.pdf

١٥- موضي بنت إبراهيم الدبيان. قياسات أداء خدمات المكتبات ومعايير تقييمه، **مجلة المعلوماتية**، تاريخ المقالة تاريخ الاستشهاد ٢٠٠٧/١١/١١ ، مسار

http://informatics.gov.sa/modules.php?name=Sections&op=printpage&artid=122

ثانيًا : المصادر الأجنبية

- **المصادر الأجنبية :**

16- 3D Object File Formats. 1996, cited 20/4/2007, cited at http://cmp.felk.cvut.cz/~pajdla/Doc/3D.formats.html

17- Assfalg, Jürgen and Del Bimbo, Alberto . Retrieval of 3D Objects by Visual Similarity, Dipartiment of Information systems, University of Firenze, Italy,2004, cited 1/4/2007, cited athttp://infovis.uni-konstanz.de/research/projects/SimSearch3D/publications/vs-icme2002.pdf

18- Bellot, Patrice. Evaluation of Document Retrieval Systems, University of Avignon , 2001, cited 20/11/2007, cited at http://www.lia.univ-avignon.fr/fich_art/146-lreceval.pdf

19- Brabrand, Claus and Møller, Anders. A Runtime System for Interactive Web Services, University of Aarhus, Denmark, 2000, cited at 1/12/2007, cited at http://www8.org/w8-papers/3b-web-doc/runtime/runtime.html

20- Breuel, Thomas. The 3D Indexing Problem, © of IDIAP, 2007, cited at 55/2007, cited at wwwwwww

21- Cacheda, Fidel and Via, Angel. Performance evaluation of Web Information Retrieval Systems and its application to e-Business, University of A Coruña, 2003, cited 25/11/2007, cited at http://www.tic.udc.es/~fidel/docs/publications/e-2002.pdf

22- Chancogne, David and Austin, Mark. Web Dieting -- Dynamic and Interactive Environment Development Tools for the Web, University of Maryland, 1998, cited at 12/12/2007, cited at www.isr.umd.eduzSz~doczSzDietzSzdietweb.pdf/web-dieting-dynamic-and.pdf

23- Chowdhury , G. G. Introduction to Modern Information Retrieval .- 2nd ed .- London: Facet publishing , 2003.

24- Cohen , Laura. Understanding the World Wide Web. Internet tutorials, 2007, cited 20/6/2007, cited at http://www.internettutorials.net/

25- Cohen, Laura. The Deep. internet tutorials, 2006, cited 20/6/2007, cita at http://www.internettutorials.net/deepweb.html

26- Database Administrator Jobs and information. UniXL , 2004, cited 2/6/2007, cited at http://www.unixl.com/dir/information_technology/databases/

27- DIG35: Metadata Standard for Digital Images. XMLCoverPages , 2002 , cited at 21/6/2004 , cited at http://xml.coverpages.org/dig35.html

28- Ditto : Privacy Policy , Ditto , 2002 , cited 10/10/2002 , http://www.ditto.com/privacy.asp

29- Dynamic URLs and Crawlers. Searchenginewatch, 2007, cited 20/12/2007, cited at

http://blog.searchenginewatch.com/blog/070815-092740

30- Find sounds : Search the Web for Sounds , Comparisonics Corporation , 2002 , cited 10/10/2002 ,

http://www.findsounds.com/help1.html

31- Gwizdka, Jacek and Chignell, Mark. Towards Information Retrieval Measures for Evaluation of Web Search Engines, University of Toronto, 1998, cuted 15/11/2007, cited at http://anarch.ie.utoronto.ca/~jacekg/pubs/webIR_eval1_99.pdf

32- Hunter , Jane and Newmarch, Jan . An Indexing, Browsing, Search and Retrieval System for Audiovisual Libraries , Uni. of Queensland , 2001 , cited 1/10/2002

33- Introduction to Databases for Web Developers : What is a Database. Extropia, 2007, cited 25/5/2007, cited at http://www.extropia.com/tutorials/sql/what_is_database.html

34- James A. O'Brien. Introduction to Information Systems: Essential for the Internet worked Enterprise. 9th.ed. NY.: McGraw-Hill, 2000, P.21

35- Jansen, Bernard J. & Sharma, Himanshu. Automated Evaluation of Search Engine Performance via Implicit User Feedback, The Pennsylvania State University, 2004, cited 15/11/2007, cited http://ist.psu.edu/faculty_pages/jjansen/academic/pubs/jansen_automated_evaluation.pdf

36- Keim , Daniel A. 3D Model Similarity Search, computer and information science research, 2006, cited 25/4/2007, cited at http://www.inf.uni-konstanz.de/index.html.en

37- Kekinen, Jaana & Jrvelin, Kalervo. EVALUATING INFORMATION RETRIEVAL SYSTEMS UNDER THE CHALLENGES OF INTERACTION AND MULTIDIMENSIONAL DYNAMIC RELEVANCE, University of Tampere, 2003, cited 1/11/2007, cited at

38- Ke-Thia Yao, Robert Neches. Dynamic Coordination of Information Management Services for Processing Dynamic Web Content, Information Sciences Institute, University of Southern California, 2005, cited 25/5/2007, cited at http://www2002.org/CDROM/refereed/613

39- Keyword Searches.[with out publisher] , 2002 , cited 10/10/2002 , cited at :http://www.concentric.net/~Rkriesel/Search/S-Keywords.htm

40- Lancaster , F. W. Vocabulary Control For Information Retrieval .- 2nd ed .- Arlington, Virginia : Information Resources Press , 1986 .- p148.

41- Luo, Lili and West, David and Marchionini, Gary. A Study of Annotations for a Consumer Health Portal , 2006, cited 20/6/2007, cited at http://ils.unc.edu/annotation/publication/luo_jcdl_2005.pdf

42- METADATA. North Carolina university, 2005, cited 20/10/2007, cited at http://www.ncecho.org/Guide/metadata.htm

43- Metadata: Image Description, Indexing and Cataloguing. tasi , 2004 , cited at 21/6/2004 , cited at http://www.tasi.ac.uk/advice/delivering/meta.html

44- metasearch engines. Clixmarketing, cited 10/10/2007, cited at www.clixmarketing.com/definitions.htm

45- metasearch.Blue Angel Technologies, Inc.2005, cited 25/9/2007, cited at http://www.BlueAngelTech.com

46- Min, Patrick , Kazhdan, Michael and Funkhouser, Thomas. A Comparison of Text and Shape Matching for Retrieval of Online 3D Models, Princeton University, 2004, cited at 20/4/2007, cited at http://citeseer.ist.psu.edu/min04comparison.html

47- Naik, sagar sudheer. 3d shape retrieval. Supervisor : parbhakaran, balakrishan , university of texas at dallas , 2006, MS. Cited at 24/4/2007, cited at http://wwwlib.umi.com/dissertations/search.

48- Nordbotten, Joan C. Multimedia Information Retrieval Systems, 2006, cited at 30/11/2007, cited at http://nordbotten.com/ADM/ADM_book/ADM-TextRetrieval.htm#R/P

49- ODLIS —Online Dictionary for Library and Information Science. Cited 10/5/2007, cited at http://lu.com/odlis/odlis_w.cfm

50- ODLIS —Online Dictionary for Library and Information Science. OP.TC.

51- O'Neill, Edward T. Trends in the Evolution of the Public Web, D-Lib Magazine, April 2003, cited 20/6/2007, cited at http://www.dlib.org/dlib/april03/lavoie/04lavoie.html

52- online DB and searching.2004, cited 15/7/2007, cited at http://web.simmons.edu/~benoit/LIS530/602slides/17online.htm

53- Palmer, Sean B. The Semantic Web: An Introduction, infomeshnet, 2001, cited 5/1/2008, cited at http://infomesh.net/2001/swintro/

54- Pirkola , ARI . Studies on Linguistic Problems and Methods in the Text Retrieval .- Tampere : University of Tampere , 1999 .- p30-31.

55- Preservation Risk Management for Web Resources. D-Lib Magazine, 2002, cited 25/5/2007, cited at http://www.dlib.org/dlib/january02/kenney/01kenney.html

56- Reitz, Joan M. directory, ODLIS — Online Dictionary for Library and Information Science, Libraries Unlimited, 2006, cited 20/8/2007, cited at http://lu.com/odlis/odlis_d.cfm# hypertext

57- Reitz, Joan M. directory, ODLIS — Online Dictionary for Library and Information Science, Libraries Unlimited, 2006, cited 20/8/2007, cited at http://lu.com/odlis/odlis_d.cfm#directory

58- Reitz, Joan M. metasearch, ODLIS — Online Dictionary for Library and Information Science, Libraries Unlimited, 2006, cited 20/9/2007, cited at http://lu.com/odlis/odlis_m.cfm# metasearch

59- Reitz, Joan M. ODLIS —Online Dictionary for Library and Information Science, 2006, cited at 1/12/2007, cited at http://lu.com/odlis/odlis_e.cfm# information retrieval

60- Reitz, Joan M. ODLIS —Online Dictionary for Library and Information Science, 2006, cited at 1/12/2007, cited at http://lu.com/odlis/odlis_e.cfm# interactive

61- Report on MP3 Searching . Search Tools consulting , 2002 , cited 2/11/2002 , http://www.searchtools.com/info/mp3-search.html.

62- Results Ranking. Net Industries, 2007, cited 20/11/2007, cited at http://ecommerce.hostip.info/pages/912/Results-Ranking.html

63- Search Portals. GILS, 2003, cited 25/11/2007, cited at http://www.gils.net/portals.html

64- Semmar, naserdine. A cross language information retrieval system, laboratory of multilanguage multimedia, 2005, cited 11/11/2007, cited at

wwwlist.cea.fr/.../gb/Alger2005_semmar_a_cross_language_information_retrieval_system.pdf

65- Smith, Alastair G. Testing the Surf: Criteria for Evaluating Internet Information Resources." The Public-Access Computer Systems Review 8, no. 3 (1997). Cited 20/5/2007, cited at http://info.lib.uh.edu/pacsrev.html.

66- Sol, Selena. Introduction to Databases for the Web, 1998, cited 10/6/2007, cited at

http://www.databasejournal.com/sqletc/article.php/26861_1431601_1

67- SUBJECT DIRECTORIES. Debbie Flanagan, 2004, cited 10/6/2007, cited at

http://www.learnwebskills.com/search/subject.html

68- Sullivan , Danny . Search Engine Optimization & Marketing Glossary, Submitawebsite, Inc , 2004 , cited at 12/6/2004 , cited at http://www.submitawebsite.com/seo_glossary.html

69- Sullivan , Danny . Search Engine Optimization & Marketing Glossary, Submitawebsite, Inc , 2004 , Visited at 12/6/2004 , http://www.submitawebsite.com/seo_glossary.html

70- Surface Web. Wikipedia : the free encyclopedia, 2007, cited 20/6/2007, cited at http://en.wikipedia.org/wiki/ Surface_Web

71- Timeless Principles of Design (Web Techniques, CMP Technology,2007, http://www.ddj.com/dept/architect/184414356?pgno=1

72- Tyner, Ross. Effective Search Strategies, staff screening, 2003, cited 20/11/2007, cited at
http://staffscreening.det.wa.edu.au/education/cmis/eval/search/search1.htm

73- Vrani, Dejan V. DESCRIPTION OF 3D-SHAPE USING A COMPLEX FUNCTION ON THE SPHERE , University of Konstanz, Department of Computer and Information Science, cited 20/4/2007, cited at http://infovis.uni-konstanz.de/research/projects/SimSearch3D/publications/vis2004.pdf

74- Watson, James. Information Models for Web Structure, published by NIVA Inc. cited 20/5/2007, Cited at http://www.writersblock.ca/spring1998/feature.htm

75- web designs. Realife web designs, 2005, cited 15/5/2007, cited at http://www.realifewebdesigns.com/web-resources/web-tips.html

76- Web protocols. Securepaynet, 2006, cited 10/5/2006, Cited at http://www.realifewebdesigns.com/web-resources/web-protocols.html

77- Website Indexing. Website-Designer, 2007, cited at 6/6/2007, cited at http://www.website-designer.net/main.php

78- What is a Search Engine? Metamend.inc , 2004 , cited at 12/6/2004 , cited at http://www.metamend.com/search-engine-map.html

79- What is a Search Engine? Metamend.inc , 2004 , Visited at 12/6/2004 , http://www.metamend.com/search-engine-map.html

80- White, Ryen W. Implicit Feedback for Interactive Information Retrieval, University of Glasgow, 2004, (doctoral thesis), cited at

https://research.microsoft.com/~ryenw/thesis/thesis-files/RyenWhiteThesis.pdf

81- White, Ryen W. Implicit Feedback for Interactive Information Retrieval, University of Glasgow, 2004, (doctoral thesis), cited at https://research.microsoft.com/~ryenw/thesis/thesis-files/RyenWhiteThesis.pdf

82- Whitson , Donna L & Amstutz , Donna D. Accessing Information in a Technological age .- Malabar, Florida : Krieger publishing company , 1997 .- p68.

83- Yang, Jun , Li, Qing and Wenyin, Liu. Retrieval of Flash™ Movies by Semantic Content: Research Issues, Generic Framework, and Future Directions, City University of Hong Kong, 2001, cited 12/12/2007, cited at www.cs.cmu.edu/~juny/Prof/papers/ieeemm-si-aesthetics.pdf

- مواقع الويب :

84- http://almashriq.hiof.no
85- http://findlaw.com
86- http://midisearch.net/
87- http://multimedia.Alltheweb.com
88- http://news.Altavista.com/
89- http://news.Google.com/
90- http://news.yahoo.com/
91- http://webmasters.abondance.com/
92- http://www.ajkids.com
93- http://www.Alltheweb.com
94- http://www.Alltheweb.com
95- http://www.Altavista.com
96- http://www.Altavista.com/mutimedia/default?qbmode=
97- http://www.business.com/
98- http://www.canadaspace.com
99- http://www.egyptsearch.com/search/

100- http://www.eserver.org/
101- http://www.Google.com
102- http://www.Google.com
103- http://www.Google.com
104- http://www.Invisibleweb.com
105- http://www.kidsClick.com
106- http://www.mamma.com
107- http://www.metaCrawler.com
108- http://www.mltavista.com
109- http://www.msnshopping.com
110- http://www.music-finder.net/
111- http://www.mwsearch.com/
112- http://www.nmmagazine.com/cgi-bin/links.cgi
113- http://www.pratique.fr/
114- http://www.search4science.com/
115- http://www.searchdesk.com/
116- http://www.searcheurope.com
117- http://www.sportec.com/
118- http://www.Webdata.com
119- http://www.yahoo.com
120- http://www.Yahooligans.com
121- http://www.yahooshopping.com
122- http://www.yupimsn.com/index.asp
123- http://wwwfroogle.com
124- www.magportal.com
125- www.Picsearch.om

الملاحق

أولاً : قائمة المصطلحات الأجنبية والمقابل العربي

المصطلح	المقابل
3D Models	نماذج ثلاثية الأبعاد
animations	الرسوم الحيوية
ARP	بروتوكول كثافة العنوان
assigned indexing	التكشيف بالتعيين
Bibliograghic databases	قواعد البيانات الببليوجرافية
bolean logoic	المنطق البولياني
Content Based Video Retrieva	الاسترجاع المبني على المحتوى
crawling	الزحف
databases search engines	محركات بحث قواعد البيانات
DBMS	نظم إدارة قواعد البيانات
DHCP	بروتوكول التعريف بالمضيف
digitization	عملية الرقمنة أو التحول إلى الشكل الرقمي
DNS	خدمة تحديد أسماء المواقع
domain name	مجال موقع الويب
DSN	اسم مصدر البيانات
Dublin Core	معيار وصف مواد الويب
Dynamic documents	صفحات الويب الديناميكية

dynamic URLs	روابط ديناميكية
dynamic web information	محتوى الويب المتغير
extended tag language	لغات الترميز الممتدة
FTP	بروتوكول نقل الملفات
hard ware	الأجهزة والمكونات المادية
header	الملف الفاتح لتسجيل الحقول
host web page	صفحة الويب المضيفة
HTML	لغة ترميز النص الفائق
HTTP	بروتوكول النص الفائق
Hypermedia	الوسائط الفائقة
ICMP	بروتوكول التحكم برسائل الإنترنت
IMAP	بروتوكول إتاحة رسائل الإنترنت
indexer	برنامج المكشف
inverted index	الكشاف المقلوب
invisble web search engines	محركات بحث الويب غير المرئية
invisible web	الويب غير المرئية
IP	بروتوكول الإنترنت
Iteraction	التفاعلية
KWAC	الكلمات الدالة المضافة للنص

KWIC	كشاف الكلمات الدالة داخل السياق
KWOC	الكلمات الدالة خارج السياق
linguistic analyzer	المحلل اللغوي
Links	الروابط بين الصفحات
MARC	الفهرسة المقروءة آليا
MARC Tages	حقول معيار الفهرسة الآلية مارك
meta search engines	محركات البحث المتعددة
Meta Search Engines	محركات البحث المتعددة
meta tag	محارف وحقول لغات الترميز
metadata	البيانات عن البيانات أو البيانات الخلفية
minimization	عملية التصغير والتخزين في البيئة الرقمية
nodes	العقد في النص الفائق
OCLC	مركز المكتبة المحسب على الخط المباشر
ODLIS	قاموس مصطلحات المكتبات والمعلومات على الخط المباشر
office systems	أنظمة إدارة سطح المكتب
Ontology Web Language	لغة الويب الوجودية
OPAC	الفهرس المتاح على الخط المباشر
page head	رأس ملف صفحة الويب
parameters	معاملات التغيير لروابط الويب

perl	لغة كتابة الواجهات العامة CGI
pixels	مقياس نقاط الكثافة
precision	معيار التحقيق
recall	معيار الاستدعاء
reformulator	برنامج البحث في النتائج
Registration	الارتباط أو الاشتراك
relevance	الصلة والارتباط بالموضوع
Reply page	صفحة الإجابة واسترجاع النتائج
resolution	كثافة المواد المصورة
Resource schema Describe Framework (RDF)	المعيار العام لوصف المصادر
response set	مجموعات النتائج المسترجعة
runtime system	نظام وقت التشغيل
script	النصوص المكودة في الويب
SDI	خدمة البث الانتقائي للمعلومات
search box	فراغ البحث
search engines	محركات البحث
Session threads	المحادثة والحوار بين النظام والمستفيد
shape-based queries	الاستفسارات المبنية على الشكل
SMTP	بروتوكول نقل البريد الإلكتروني

برامج الحاسب الآلي	soft ware
عمل برنامج العنكبوت	spidering
قوائم التوقف في التكشيف الإلكتروني	Stop List
الويب السطحية " المرئية"	surface web
بروتوكول التحكم بالنقل / بروتوكول الإنترنت	TCP/IP
بروتوكول الاتصال النقل عن بعد	TELNET
بنية المواد المعلوماتية	texture
المكانز	Thesaurus
اللقطات	thumbnails
المصطلح الموحد في التكشيف	uni term
الحقول المتغيرة في المحتوى	variable fields
متصفح الويب	Web browser
دليل الويب	web directory
الإبحار عبر الويب	web navigation
خادمات الويب	web servers
لغة الترميز الممتدة	XML
بوابة البحث ياهو	Yahoo
معيار استرجاع المعلومات	Z39.50

ثانيًا : قائمة المصطلحات العربية والمقابل الأجنبي

المصطلح	المقابل
الإبحار عبر الويب	web navigation
الأجهزة والمكونات المادية	hard ware
أنظمة إدارة سطح المكتب	office systems
الارتباط أو الاشتراك	Registration
الاسترجاع المبني على المحتوى	Content Based Video Retrieva
الاستفسارات المبنية على الشكل	shape-based queries
اسم مصدر البيانات	DSN
برامج الحاسب الآلي	soft ware
برنامج البحث في النتائج	reformulator
برنامج المكشف	indexer
بروتوكول إتاحة رسائل الإنترنت	IMAP
بروتوكول الإنترنت	IP
بروتوكول الاتصال النقل عن بعد	TELNET
بروتوكول التحكم بالنقل / بروتوكول الإنترنت	TCP/IP
بروتوكول التحكم برسائل الإنترنت	ICMP
بروتوكول التعريف بالمضيف	DHCP
بروتوكول النص الفائق	HTTP
بروتوكول كثافة العنوان	ARP

بروتوكول نقل البريد الإلكتروني	SMTP
بروتوكول نقل الملفات	FTP
بنية المواد المعلوماتية	texture
بوابة البحث ياهو	Yahoo
البيانات عن البيانات أو البيانات الخلفية	metadata
التفاعلية	Iteraction
التكشيف بالتعيين	assigned indexing
الحقول المتغيرة في المحتوى	variable fields
حقول معيار الفهرسة الآلية مارك	MARC Tages
خادمات الويب	web servers
خدمة البث الانتقائي للمعلومات	SDI
خدمة تحديد أسماء المواقع	DNS
دليل الويب	web directory
رأس ملف صفحة الويب	page head
الرسوم الحيوية	animations
الروابط بين الصفحات	Links
روابط ديناميكية	dynamic URLs
الزحف	crawling
صفحات الويب الديناميكية	Dynamic documents
صفحة الإجابة واسترجاع النتائج	Reply page
صفحة الويب المضيفة	host web page

الصلة والارتباط بالموضوع	relevance
العقد في النص الفائق	nodes
عمل برنامج العنكبوت	spidering
عملية التصغير والتخزين في البيئة الرقمية	minimization
عملية الرقمنة أو التحول إلى الشكل الرقمي	digitization
فراغ البحث	search box
الفهرس المتاح على الخط المباشر	OPAC
الفهرسة المقروءة آليا	MARC
قاموس مصطلحات المكتبات والمعلومات على الخط المباشر	ODLIS
قوائم التوقف في التكشيف الإلكتروني	Stop List
قواعد البيانات الببليوجرافية	Bibliograghic databases
كثافة المواد المصورة	resolution
كشاف الكلمات الدالة داخل السياق	KWIC
الكشاف المقلوب	inverted index
الكلمات الدالة المضافة للنص	KWAC
الكلمات الدالة خارج السياق	KWOC
لغات الترميز الممتدة	extended tag language
لغة الترميز الممتدة	XML
لغة الويب الوجودية	Ontology Web Language
لغة ترميز النص الفائق	HTML

perl	لغة كتابة الواجهات العامة CGI
thumbnails	اللقطات
Web browser	متصفح الويب
domain name	مجال موقع الويب
response set	مجموعات النتائج المسترجعة
Session threads	المحادثة والحوار بين النظام والمستفيد
meta tag	محارف وحقول لغات الترميز
dynamic web information	محتوى الويب المتغير
search engines	محركات البحث
meta search engines	محركات البحث المتعددة
Meta Search Engines	محركات البحث المتعددة
invisble web search engines	محركات بحث الويب غير المرئية
databases search engines	محركات بحث قواعد البيانات
linguistic analyzer	المحلل اللغوي
OCLC	مركز المكتبة المحسب على الخط المباشر
uni term	المصطلح الموحد في التكشيف
parameters	معاملات التغيير لروابط الويب
Z39.50	معيار استرجاع المعلومات
recall	معيار الاستدعاء
precision	معيار التحقيق
Resource schema Describe Framework (RDF)	المعيار العام لوصف المصادر

معيار وصف مواد الويب	Dublin Core
مقياس نقاط الكثافة	pixels
المكانز	Thesaurus
الملف الفاتح لتسجيل الحقول	header
المنطق البولياني	bolean logoic
النصوص المكودة في الويب	script
نظام وقت التشغيل	runtime system
نظم إدارة قواعد البيانات	DBMS
نماذج ثلاثية الأبعاد	3D Models
الوسائط الفائقة	Hypermedia
الويب السطحية " المرئية"	surface web
الويب غير المرئية	invisible web

الكتـــاب :

إن مصطلح نظم استرجاع المعلومات يدل على مفهوم عام ينسحب على مختلف مرافق المعلومات على اختلاف أنواعها ودرجاتها. غير أن الإنتاج الفكري الأجنبي والعربي دائماً ما يتحدث عن نظم استرجاع المعلومات من منظور واحد فقط، ألا وهو استراتيجيات ولغات بحث المعلومات.

استعرض الباحث الإنتاج الفكري في قاعدة مستخلصات علم المكتبات والمعلومات LISA الخاص بمفهوم نظم استرجاع المعلومات فلم يجد من الدراسات ما أعطى لهذا المفهوم حقه كاملاً. ولم يرد مفهوم نظم استرجاع المعلومات في قاموس مصطلحات المكتبات وعلم المعلومات على الخط المباشر ODLIS بشكل مخصص، حيث عرف ODLIS مفهوم استرجاع المعلومات على أنه: العملية أو الإجراء أو الطريقة التي تنتقي مجموعة من المعلومات المسجلة لاستدعائها بحسب مجموعة من المصطلحات البحثية التي تعبر عن حاجة موضوعية محددة، وهذه العملية قد تكون تقليدية كما يتم في استخدام الفهارس البطاقية في المكتبات أو قد تكون إلكترونية كما يتم في الفهارس على الخط المباشر OPAC أو مراصد البيانات الببليوجرافية "Bibliograghic databases" ومن ثم فإن هذا المفهوم يركز على الجانب الإجرائي والأدائي الذي تتم عليه عملية استرجاع المعلومات سواء أكان ذلك في البيئة التقليدية أو الرقمية.

إن ظهور الإنترنت والويب وأدوات بحث مصادرها الرقمية المختلفة الأشكال والمتباينة الوسائط والمتعددة الموضوعات قد دفع الباحث إلى النظر لما أحدثته البيئة الرقمية المتمثلة في الإنترنت وخاصة الويب في تغيير إدارة وعمل ومهام نظم استرجاع المعلومات. فلقد ظلت نظم استرجاع المعلومات حتى الآلية منها تسير سيراً منطقياً في إجراءات تشغيلها ومعالجتها وتنظيمها للمعلومات ثم استراتيجيات بحثها مرة أخرى، غير أن الويب قد أحدثت تغييرات عدة في ملامح نظم استرجاع المعلومات، وإن جاز هنا أن نطلق عليها نظام النظم أو نظام استرجاع المعلومات، وبدت هذه التغييرات ظاهرة ومعالجة واسترجاع المعلومات الرقمية. وقد أراد الباحث من خلال هذا البحث دراسة فضاء الويب ومعايشة عناصر بنيتها وملامح تشغيلها كنظام كوني لاسترجاع المعلومات. والهدف من ذلك كله هو تقديم رؤية جديدة للتعامل مع الويب بيسر وتحقيق الكفاية المرجوة في استرجاع المصادر الرقمية على الويب.

المـؤلـف :

سيد ربيع سيد إبراهيم .

- ليسانس آداب جامعة القاهرة – فرع بني سويف عام ٢٠٠٠ م.
- ماجستير في الآداب.
- مسجل لنيل درجة الدكتوراة.
- اجتاز عدداً من الدورات العلمية .
- له كتاب منشور تحت عنوان محركات بحث الصور الثابتة.
- له أبحاث علمية منشورة .
- يعمل الآن مدرساً مساعداً بقسم المكتبات والوثائق – كلية الآداب – جامعة بني سويف.

ردمك : ٤- ٣٣٤ -٠٠- ٩٩٦٠- ٩٧٨

